

FRASEOLOGIA EDISTYNEEN SUOMENOPPIJAN KOMPASTUSKIVENÄ
Epätyypillisyyksiä C-tason oppijansuomessa

Suomen kielen
pro gradu -tutkielma
Oulun yliopisto
toukokuu 2021

Ying Xi Cheng

SISÄLLYS

1. JOHDANTO	1
1.1. Tutkimusaihe ja keskeiset käsitteet	1
1.2. Tutkimuskysymykset	2
2. TUTKIMUSAINEISTO JA -MENETELMÄT	4
2.1. Korpus ja korpuslingvistiikka	4
2.1.1. Korpustutkimuksen eri lähestymistavat	5
2.1.2. Oppijankielen korpustutkimus	7
2.2. Tutkimusaineisto	10
2.2.1. Kansainvälinen oppijansuomen korpus	10
2.2.2. Natiivisuomen vertailuaineisto	11
2.3. Tutkimusmenetelmät	12
2.3.1. Epätyypillisyyksien valinta	12
2.3.2. Epätyypillisyyksien käsittely ja luokittelu	13
2.4. Tutkimukseen liittyvät ongelmat	14
2.4.1. Kvantitatiivisten menetelmien vähäisyys	14
2.4.2. Intuition puutteita	15
3. TEORIA JA TAUSTA	18
3.1. Kontekstuaalinen semantiikka	18
3.1.1. Käsitteen tausta ja määritelmä	18
3.1.2. Tilannekonteksti, konteksti ja koteksti	19
3.1.3. Syntagmaattiset ja paradigmaattiset suhteet	20
3.1.4. Vapaan valinnan periaate ja idiomiperiaate	20
3.1.5. Leksikaalinen ja morfologinen priming	22
3.2. Fraseologia ja fraseologinen yksikkö	23
3.2.1. Käsitteen tausta ja määritelmä	23
3.2.2. Fraseologinen yksikkö	25
3.2.3. Keskeiset fraseologiset yksiköt ja käsitteet	28
3.2.3.1. Kollokaatio	29
3.2.3.2. Kolligaatio	31
3.2.3.3. Semanttinen preferenssi	32

3.2.3.4. Semanttinen prosodia	32
3.3. Oppijankieli ja edistynyt kielitaito	34
3.3.1. Oppijankielen käsite ja luonne	34
3.3.2. Virheet ja epätyypillisyydet	36
3.3.3. Kielitaito ja Eurooppalainen viitekehys	38
3.3.4. Edistyneen kielenoppijan fraseologiset haasteet	39
3.4. Aiempi tutkimus	40
4. ANALYYSI	42
4.1. Epätyypillisyyksien valintaperusteet, luokittelu ja käsittely	42
4.2. Sananvalinnan epätyypillisyydet	44
4.2.1. Yksittäiset lähimerkityksiset sanat	44
4.2.2. Intensiteettipartikkelit	51
4.2.3. Muut kollokaatiot ja leksikaaliset myötäesiintymät	56
4.2.4. Semanttinen preferenssi ja semanttinen prosodia	60
4.2.5. Uudismuodosteet	63
4.2.6. Sanakirjan sudenkuopat ja transfer-tapaukset	65
4.2.7. Virheelliset nomininvalinnat	69
4.2.8. Virheelliset verbinvalinnat	70
4.3. Morfologiset epätyypillisyydet	73
4.4. Lauseketason epätyypillisyydet	77
4.4.1. Substantiivilausekkeet	78
4.4.2. Adjektiivi- ja partisiippilausekkeet	80
4.4.3. Adpositiolausekkeet	81
4.4.4. Infinitiivilausekkeet	84
4.5. Lausetason epätyypillisyydet	86
4.5.1. Verbin valenssi	87
4.5.2. Verbi ja sen täydennysympäristö	88
4.5.3. Lausetyypit	89
4.5.4. <i>Tulla joksikin</i> -rakenne	91
4.5.5. Vierasperäiset lauserakenteet	93
4.5.6. Kiteytyneet rakenteet	94
4.6. Yhteenvedo ja tulosten merkitys oppijansuomen opetuksen kannalta	96

5. PÄÄTÄNTÖ	100
LÄHTEET JA LYHENTEET	104
LIITE	

1. JOHDANTO

1.1. Tutkimusaihe ja keskeiset käsitteet

Muutin Suomeen ummikkona noin kaksitoista vuotta sitten. Kun olin vielä kehittämässä sitä suomen kielen taitoa, joka nyt mahdollistaa tämän tutkielman kirjoittamisen, yksi asia jäi askarruttamaan: pelotteluista ja ponnisteluista huolimatta en kokenut sanaston ja kieliopin omaksumista erityisen hankalaksi. Sen sijaan todellinen vaikeus on ollut kielen tuottamisessa niin, että se kävisi äidinkielen kielenkäyttäjän tuotoksesta. Olen käyttänyt suomea päivittäin miltei ongelmitta jo useita vuosia. Silti usein puheessani ja kirjoituksissani ilmenee sellaisia kohtia, jotka saattavat särähtää äidinkielen kielenkäyttäjän korvaan ja paljastaa ulkomaalaistaustani. Nämä niin sanotut epäsuomalaisuudet ovat kiinnostaneet minua jo pitkään.

Tässä pro gradu -tutkielmassani tarkastelen ja kartoitan niitä fraseologisia epätyypillisyyksiä, joita vielä esiintyy edistyneiden suomenoppijoiden kirjallisissa tuotoksissa. **Fraseologisuudella** tarkoitan erilaisten fraasien, sanontojen, idiomien ja keskustelurutiinien lisäksi myös muita vakiintuneita leksikaalisia yhdistelmiä, jotka muodostavat huomattavasti laajemman sanastollis-kieliopillis-semanttisen ja tekstuaalisen ilmiön (Granger 1998a; Jantunen & Brunni 2012: 75). Termillä **suomenoppija** (yleisemmällä tasolla **kielenoppija**) viitataan ihmiseen, joka on oppinut suomea joko vieraana kielenä kieliyhteisön ulkopuolella tai toisena kielenä kieliyhteisön parissa. Fraseologisella epätyypillisyydellä tarkoitan siis suomenoppijan kielentuotosta, joka poikkeaa jollain lailla vastaavasta äidinkielen kielenkäyttäjän eli **natiivin** kielentuotoksesta. Edistyneellä suomenoppijalla taas tarkoitan henkilöä, joka kykenee tuottamaan Eurooppalaisen viitekehyksen (EVK) mukaan C1–C2-tasoiseksi arvioituja tekstejä. Vastaavasti edistynyt oppijansuomi koostuu kielentuotoksesta, joka on arvioitu C1–C2-tasoiseksi.

Yksi tutkimukseni tavoitteista on selvittää, miten edistyneen suomenoppijan kielenkäyttö saattaa erota natiivin kielenkäytöstä niin, että muuten kieliopillisesti virheettömästä ja semanttisesti ymmärrettävästä tekstistä paljastuu jokin ”epäsuomalaisen kuuloinen” piirre. Vaikka tutkimukseni kuuluukin korpuslingvistiikan piiriin, käsittelen aineistoani lähinnä kvalitatiivisesti. Oppijansuomen aineistoni muodostuu Oulun yliopiston Kansainvälisen oppijansuomen korpuksen (ICLFI) C1–C2-tasoisista teksteistä. Käyn kyseiset tekstit

manuaalisesti läpi ja poimin niistä tapauksia, jotka tuntuvat kieli-intuitioni perusteella poikkeavan vastaavasta natiivisuomen tuotoksesta. Tarkistaakseni tapausten poikkeavuuden ja havainnollistaakseni näiden kahden kielimuodon eroja hyödynnän natiivisuomen korpusta vertailuaineistona. Konkordanssihaun avulla pystyn vertailemaan fraseologisesti tyypillisten ja epätyypillisten tapausten esiintymiskertoja toisiinsa ja siten testaamaan alustavien havaintojeni paikkansapitävyyttä.

1.2. Tutkimuskysymykset

Tutkimukseni pohjautuu aiemmin tekemääni samanaiheiseen kandidaatintutkielmaan (Cheng 2014). Siinä valitsin omaan ja vähintään yhden äidinkielen kielenkäyttäjän kieli-intuition nojautuen 18 fraseologisesti epätyypillistä tapausta ICLFIstä tarkasteltavaksi ja laadin niille oman alustavan luokituksen. Tämän pro gradu -tutkielman kartoitus on paljon laajempi ja perusteellisempi, joskin sekään ei ole tyhjentävä. Hyödynnän kartoituksessani hiljattain ilmestynyttä uutta tutkimustietoa, erityisesti Kuuluvaisen (2015) luokittelutapaa ICLFI:n fraseologisiksi virheiksi koodatulle virheannotoidulle aineistolle.

Grangerin (1998b: 15) mukaan tutkimukset voivat olla hypoteesiin perustuvia (*hypothesis-based*) tai hypoteesia etsiviä (*hypothesis-finding*). Hän määrittelee hypoteesiin perustuvan lähestymistavan sellaiseksi, jossa tutkija muodostaa hypoteesinsa etukäteen alan tutkimuskirjallisuuden pohjalta ja testaa sitä korpusaineistoa hyödyntäen. Scholfield (1995: 24) taas luonnehtii hypoteesia etsivää lähestymistapaa sellaiseksi, jossa tutkija ensin kerää ja käsittelee kaikki aineistonsa ja vasta sen jälkeen tarkastelee sitä mahdollisten johtopäätösten löytämiseksi. Yhdistän tutkimuksessani molemmat lähestymistavat, sillä pyrin sekä vahvistamaan olemassa olevia tutkimuslöydöksiä että löytämään uusia havaintoja edistyneen oppijansuomen luonteesta.

Tutkimuskysymykseni ovat seuraavat:

1. Millaisia fraseologisia epätyypillisyyksiä yhä esiintyy edistyneiden suomenoppijoiden kirjallisissa tuotoksissa?
2. Miten nämä epätyypillisyydet ovat luokiteltavissa varsinkin uuden tutkimuksen valossa?

3. Onko tutkimissani suomenoppijoiden tuotoksissa sellaisia fraseologisen epätayppillisuuden piirteitä, jotka olisivat yleistettävissä?
4. Voiko tutkimukseni tuloksia hyödyntää oppijasuomen opetuksessa ja jos voi, millä tavalla?

2. TUTKIMUSAINETO JA -MENETELMÄT

2.1. Korpus ja korpuslingvistiikka

Korpus on jonkin luonnollisesti esiintyvän (*naturally-occurring*) kielen sähköinen tekstikokoelma, jota voidaan käyttää kielitieteelliseen analyysiin. Korpuksen tekstit on yleensä valittu edustamaan tietynlaista kielivarianttia, ja ne on kerätty tarkkoja kriteereitä noudattaen ja tiettyä tarkoitusta varten. (Sinclair 1991: 171; Biber, Conrad & Reppen 1998: 12; Tognini-Bonelli 2001: 2; Hunston 2006: 234.)

Korpuslingvistiikan määrittely taas ei ole ongelmaton, sillä alalla vallitsee useita näkemyksiä siitä, tulisiko se luokitella teoriaksi vai metodiksi (Tognini-Bonelli 2001: 1). Noudatan tässä tutkielmassa muun muassa Grangerin (2002: 4), Hunstonin (2006: 234, 244), Griesin (2009: 1) sekä Hardien ja McEneryn (2010: 385) ajatusta siitä, että korpuslingvistiikka on kielitieteellinen metodologia, joka pohjautuu korpuksen eli autenttista teksteistä koostuvan sähköisen kieliaineiston hyödyntämiseen.

Hunstonin (2006: 234) mukaan korpuslingvistiikkaa voidaan pitää pitkälle kehittyneenä menetelmänä, jonka avulla etsitään vastauksia kieleen liittyviin kysymyksiin. Hän lisää, että korpuksia hyödyntämällä pystytään testaamaan erilaisia hypoteeseja sekä tuomaan kvantitatiivinen ulottuvuus kielitieteellisiin tutkimuksiin. Biber, Conrad ja Reppen (1998: 4) kiteyttävät korpusmenetelmän keskeisimmät ominaisuudet seuraaviin kohtiin: korpusmenetelmässä analysoidaan empiirisesti kielen käyttötapoja (*patterns of use*), käytetään korpuksia analyysin pohjana, hyödynnetään laajamittaisesti korpusaineistoja käsitteleviä tietokoneohjelmia sekä nojaututaan niin kvantitatiivisiin kuin kvalitatiivisiinkin analyysimenetelmiin.

Korpusmenetelmän etuja ovat muun muassa tutkimusaineistojen laajuus sekä tietokoneohjelmilla toteutettavat automaattiset analyysitoiminnot, jotka mahdollistavat aiemmin saavuttamattomia, hyvinkin monimutkaisia kvantitatiivisia analyyseja. Nykyaikaisella tehokkaiden tietokoneohjelmien ja laajojen korpusaineistojen yhdistelmällä pystytään tuomaan esille sellaista todelliseen kielenkäyttöön perustuvaa tietoa, joka ennen olisi ollut tutkijoiden saavuttamattomissa. (Granger 2002: 4; Jantunen 2012b: 360.) Kun osa tietokannan ylläpitotehtävistä ja analyysitoiminnoista siirtyy tietokoneohjelmien hoidetta-

vaksi, tutkija on vapaa keskittymään kielitieteellisten päätelmiensä tekemiseen (Biber, Conrad & Reppen 1998: 4).

Korpuksen toinen etu on raaka-aineiston **annotoinnin** mahdollisuus (ks. esim. Granger 2002: 13–14). Vaikka Heikkisen, Lounelan ja Voutilaisen (2012: 374) mukaan käsitykset annotaatiosta vaihtelevat, se viittaa perinteisesti jonkinlaista lingvististä analyysia edustavaan tietoon, joka lisätään raakaan kieliaineistoon esimerkiksi koodin muodossa (Leech 1993: 275; Gries 2009: 9–10). Yksi esimerkki annotoinnista on **lemmatisointi**, jossa vaihtelevien sananmuotojen taustalta hahmotetaan sanan perusmuoto eli **lemma** (Gries 2009: 10; Heikkinen, Lounela & Voutilainen 2012: 375). Toinen erityisesti oppijankielen korpuksille ominainen annotointitapa on **virheannotointi** (ks. luku 2.1.1.).

2.1.1. Korpustutkimuksen eri lähestymistavat

Korpus tarjoaa empiirisen ulottuvuuden perinteisille introspektioon pohjautuville havainnoille (Arppe 2008: 2). Kun aiemmin lingvistiset teoriat muodostettiin tutkijan runsaan kielitiedon ja pohdiskelun pohjalta ja testattiin pätevien kielenkäyttäjien kieli-intuitioon tukeutuen, korpuksen avulla pystytään testaamaan kieliteorioiden paikkansapitävyyttä empiirisin keinoin, havainnollistamaan niitä esimerkeillä ja jopa luomaan niitä (Tognini-Bonelli 2001: 65). Korpus tarjoaa kolmenlaista tietoa: empiiristä, laadullista tietoa hypoteesien tukemiseksi, frekvenssitietoa kieliainesten taajuudesta sekä metatietoa, joka mahdollistaa erilaisten tekstien ja puhujaryhmien vertailun. Korpusdata ei kuitenkaan voi korvata introspektiota, kokeellista dataa eikä kenttättyötä. (Lüdeling & Kytö 2008: viii–ix.)

Hardie ja McEneryn (2010: 384–385) mukaan korpuslingvistiikassa vallitsevat lähestymistavat voidaan jakaa karkeasti kahteen päälinjaan. Näitä päälinjoja Tognini-Bonelli (2001) kutsuu **korpuspohjaiseksi** (*corpus-based*) ja **korpusvetoiseksi** (*corpus-driven*) tutkimukseksi. Jantunen (2009a: 101–102) lisää joukkoon vielä kolmannen lähestymistavan, **aineistoesimerkein tuetun** (*corpus-illustrated*) tutkimuksen (ks. Tummers, Heylen & Geeraerts 2005: 234–238). Kyseinen lähestymistapa antaa Jantusen (mp.) mukaan vähiten painoarvoa aineistolle, sillä siinä aineisto toimii lähinnä materiaalipankkina, josta etsitään sopivia – yleensä tutkijan kieli-intuitiota tukevia tai täydentäviä – esimerkkejä tietyn ilmiön kuvaamiseksi.

Aineistoesimerkein tuettuun tutkimukseen verrattuna korpuksen rooli on selvästi suurempi korpuspohjaisessa tutkimuksessa, sillä aineisto ei toimi pelkkänä esimerkkivaranona, vaan sitä myös analysoidaan laajamittaisesti (Jantunen 2009a: 102). Korpuspohjaisessa tutkimuksessa korpuksen roolina on pääasiassa ennalta muodostettujen havaintojen, teorioiden tai hypoteesien testaaminen, selittäminen tai havainnollistaminen esimerkein. Nämä lähtökohtana toimivat havainnot, teoriat ja hypoteesit voivat perustua tutkijan intuitioon tai aiempiin tutkimuksiin, ja niiden pohjalta syntyviä tutkimuskysymyksiä tarkastellaan korpusaineiston avulla. Korpuspohjaisessa tutkimuksessa aineistoa analysoidaan olemassa olevien kielellisten kategorioiden ja kuvausten puitteissa. Sen avulla näitä kategorioita voidaan muokkailla ja parannella, mutta koska kategorioiden ei voida sanoa syntyneen suoraan aineistosta, niitä ei myöskään pystytä kumoamaan pelkän aineiston perusteella. (Tognini-Bonelli 2001: 65–66.)

Jantunen (2009a: 103) mukaan kunkin kolmen lähestymistavan joukosta aineiston merkitys on suurin korpusvetoisessa tutkimuksessa. Toisin kuin korpuspohjaisessa lähestymistavassa, korpusvetoisen lähestymistavan tavoitteena ei ole pelkästään testata, päivittää tai havainnollistaa aiempia havaintoja, teorioita ja hypoteeseja. Sen sijaan aineistoa tarkastellaan kokonaisvaltaisesti eikä pelkästään esimerkkivaranona tai olemassa olevan järjestelmän mahdollisena jatko-osana. Aineistolle annettu painoarvo hypoteesien luomisessa, teorioiden kehittämisessä ja kielen ilmiöiden kategorisoinnissa on sen verran suuri, että tehdyt havainnot ja johtopäätökset tulevat suoraan korpuksen todistusaineistosta ja ovat yhdenmukaisia sen kanssa. Usein päädytään myös johtopäätöksiin, joita ei ole mahdollista tehdä muuten kuin korpuksen todistusaineistoa tarkastelemalla. Toistuvuus ja frekvenssit ovat keskeisessä asemassa; itse asiassa niitä pidetään usein kielellisten kategorioiden perustana. Myös jonkin kielellisen ilmiön puuttumista aineistosta pidetään mahdollisesti merkittävänä teorian kannalta. (Tognini-Bonelli 2001: 84.)

Oman tutkimukseni määrittelemisen näiden kolmen lähestymistavan puitteissa on hienoa hankalaa, sillä Kuuluvaista (2015) lukuun ottamatta tällaisia laajempia oppijansuomen yleiskartoituksia ei ole tietääkseni vielä tehty. Tutkimukseni muistuttanee eniten korpusvetoista tutkimusta, sillä annan korpusaineistolleni suuren painoarvon, ja tekemäni hypoteesit ja päätelmät syntyvät suoraan siitä. Kuitenkaan korpusvetoiselle tutkimukselle ominaiset toistuvuus- ja frekvenssianalyysit eivät kuulu tutkimusmenetelmäni, vaan keskityn enimmäkseen aineistoni laadulliseen tarkasteluun.

2.1.2. Oppijankielen korpustutkimus

Oppijankielen korpustutkimus (*learner corpus research*) on suhteellisen uusi korpuslingvistiikan osa-alue. Saatuaan jalansijaa korpuslingvistiikassa 1980-luvun loppupuolella sen suosio on kasvanut huomattavasti 1990- ja 2000-luvulla, ja se on toiminut siltana ennen erillisten korpuslingvistiikan ja oppijankielen tutkimusalojen välillä. (Granger 2002: 4–5; 2008: 259.) Suosion kasvun myötä on onnistuttu siirtymään eteenpäin siitä ajattelutavasta, että ainoastaan natiivien kielentuotoksia ovat tutkimuksen arvoisia (Mauranen 2011: 156). Alkuperäisen oppijanenglannin lisäksi on myös alettu keräämään oppijankielen aineistoja muistakin kielistä, ja nykyään ala on kasvanut maailmanlaajuiseksi (Jantunen & Pirkola 2015: 88).

Oppijankielen korpustutkimuksen avulla pyritään luomaan entistä parempi kuvaus oppijankielestä, mikä puolestaan auttaa kehittämään sekä toisen kielen omaksumista (*Second Language Acquisition*) tutkivaa kenttää että vieraan kielen opetuksen (*Foreign Language Teaching*) ja opiskelun menetelmiä. Kun aiemmin toisen kielen omaksumisen tutkimuksissa suosittiin kokeellista ja introspektioon perustuvaa tietoa todellisten kielenkäyttöesimerkkien sijaan, nykyään on teknologian ansiosta mahdollista kerätä ja analysoida isoja määriä oppijankielen aineistoa tietokoneohjelmien avulla. Niinpä huomio on alkanut kääntyä entistä korpuspainotteisempaan suuntaan molemmilla tutkimuskentillä. (Granger 2002: 5–6; 2008: 259.)

Oppijankielen korpusten avulla voidaan esimerkiksi vertailla kielenoppijoiden tuotoksia natiivien tuotoksiin ja tunnistaa virheitä tai epätyypillisyyksiä oppijankielessä. Yleensä oppijankielestä havaitaan tietynlainen kielenpiirre, joka yli- tai aliedustuu rinnastettavissa olevaan natiivikielen verrattuna. Sen jälkeen tätä tietoa käytetään kvalitatiivisemmän analyysin pohjana. (Hunston 2006: 246.) Tutkimuksen kannalta tärkeää on se, että oppijankielen korpusten avulla voidaan tutkia myös edistynyttä oppijankieltä paremmin ja näin saada parempi kuva siitä, mikä erottaa edistyneitä kielenoppijoita äidinkielisistä kielenkäyttäjistä ja mistä nämä erot johtuvat (Ivaska 2015: 3–4).

Oppijankielen korpus on yksinkertaisimmillaan sähköinen tekstikokoelma, joka koostuu kielenoppijoiden tuottamista teksteistä (Nesselhauf 2004b: 125; Hunston 2006: 246; Granger 2008: 259). Ne lasketaan usein erikoiskorpusten joukkoon (ks. esim. Jantunen

2011: 87). Erikoiskorpukset eroavat Tognini-Bonellin (2001: 8) mukaan yleiskorpuksista siinä, että ne sisältävät huomattavassa määrin tavallisesta kielestä poikkeavia ominaispiirteitä, kun taas yleiskorpukset kuvaavat laajasti tavallista, autenttista kieltä.

Granger (2002: 7–10) korostaa, että oppijankielen korpuksen on koostuttava yhtenäisistä teksteistä, jotka sisältävät sekä oikein että virheellisesti tuotettua kieltä; pelkästään yksittäisten sanojen tai lauseiden kokoelmaa ei siis voi kutsua korpukseksi. Lisäksi mm. Sinclair (1996), Granger (mp.) ja Nesselhauf (2004b: 130) painottavat tiukkojen kriteereiden noudattamisen tärkeyttä oppijankielen korpuksen kokoamisessa: koska oppijankielessä esiintyy niin paljon variaatiota, on taustamuuttujien huomioonottaminen ensiarvoisen tärkeää luotettavien ja tarkkojen tutkimustulosten saamiseksi. Myös Jantunen (2011: 93) on sitä mieltä, että mitä monipuolisempaa tekstien taustamuuttujien keräys ja mitä parempaa niiden dokumentointi ovat, sitä parempia tutkimuskäyttömahdollisuuksia aineistolla on.

Oppijankielen korpus voi olla joko pelkistä teksteistä koostuva raakakorpus tai annotoitu korpus, jonka teksteihin on lisätty lingvististä tai tekstuaalista informaatiota (Granger 2002: 10). Yleisin annotaatio oppijankielen korpuksissa on virheannotaatio (Barlow 2005: 341), mikä johtunee osittain alan virheanalyysipainotteisuudesta (Jantunen 2011: 98). Virheannotaatio on tosin osoittautunut ongelmalliseksi useista syistä. Esimerkiksi oppijankielen luonne saattaa aiheuttaa monitulkintaisuuksia virhetapausten tulkinnessa (Jantunen & Piltonen 2009: 450), ja lisäksi virheiden koodaus on pitkälti manuaalista ja siten työlästä ja aikaa vievää tarkistusvaiheineen (Dagneaux, Denness & Grager 1998: 167; Jantunen 2011: 98; Jantunen, Brunni, Lehto & Airaksinen 2014: 71).

Oppijankielen korpuksilla on muiden korpusten tavoin tietynlaiset ominaisuudet, joita Granger (2002: 10–11) luonnehtii seuraavasti: oppijankielen korpukset ovat pääosin yksikielisiä ja koostuvat pikemminkin yleis- kuin erikoiskielestä. Ne tapaavat olla synkronisia eli tietynaikaista kielenkäyttöä kuvailevia. Suullisen kieliaineiston keräämiseen liittyvien hankaluuksien takia kirjoitetusta kielestä koostuvia oppijankielen korpuksia on enemmän kuin puhutun kielen korpuksia. Lisäksi oppijankielen korpukseen tallennetaan tiedot kielenoppijoita ja kielitehtäviä koskevista muuttujista, joiden perusteella tutkijat voivat koostaa erilaisia osakorpuksia omia tutkimustarkoituksiinsa varten. (Mp.)

Yleensä oletetaan, että korpukseen tallennetaan vain luonnollisesti esiintyvää, autenttista kieltä (Tognini-Bonelli 2001: 2). Muun muassa Granger (2002: 8) on kuitenkin huomauttanut, että autenttisuuden tarkka määrittelemineen on vaikeaa, kun kyseessä on oppijankielen korpusaineisto. Hän toteaa, että autenttisinkin kerätty oppijankielen tuotos on harvoin yhtä autenttista kuin natiivin tuottama kieli. Tämä pätee erityisesti silloin, kun aineisto on kerätty osana muodollista kielen opiskelua, jolloin tietynasteista keinotekoisuutta on aina mukana. On siksi määriteltävä autenttisuuden käsite uudelleen oppijankielen kontekstissa ja sisällytettävä siihen sellaisiakin tapauksia, joissa oppijankieltä tuotetaan pyynnöstä erilaisissa tehtävissä. (Mp.)

Oppijansuomen korpustutkimus käynnistyi Suomessa 2000-luvun alussa, ja sen tienviitana on toiminut oppijanenglannin korpustutkimus (Jantunen & Brunni 2012: 72). Tähän mennessä oppijansuomea kuvaavia korpuksia on yhteensä seitsemän: Oulun yliopiston Kansainvälinen oppijansuomen korpus (ICLFI), Turun yliopiston Edistyneiden suomenoppijoiden korpus (LAS2), Jyväskylän yliopiston Yleisten kielitutkintojen korpus (YKI-korpus), Cefling-korpus, Topling-korpus ja Dialuki-korpus sekä Helsingin ja Tallinnan yliopiston yhteistyöhankkeena kokoama Long Second -korpus. Näistä korpuksista ainoastaan ICLFI koostuu suomi vieraana kielenä -aineistosta; muiden korpusten sisältämät kieliaineistot edustavat suomi toisena kielenä -varianttia. Korpusten välillä esiintyy paljon variaatiota tuottamis- ja oppimisympäristön, kielitaitotason, tekstilajien, aineiston pitkittäisyyden sekä oppijoiden ensikielten ja iän osalta. (Jantunen & Pirkola 2015.)

Yleisesti ottaen korpuksien ovat vaikuttaneet kielen opetukseen useammalla tavalla. Esimerkiksi korpustutkimuksen tuloksia on laajasti hyödynnetty kielenoppimismateriaalien kehittämisessä, kielenoppijoita on alettu kannustaa tutustumaan korpuksiin ja korpusmenetelmä on otettu käyttöön oppijankielen tutkimuksissa. (Hunston 2006: 245.) Lisäksi korpustutkimuksen tuloksia voidaan käyttää hyväksi kohdennettujen, todelliseen kielenkäyttöön perustuvien opetustyökalujen ja kielitestien suunnittelussa (Biber, Conrad & Reppen 1998: 12). Tästä huolimatta korpusmenetelmän juurtuminen oppijansuomen tutkimukseen ja opetukseen lieenee vasta alullaan. Esimerkiksi Jantunen (2009: 374) on todennut, että tietokoneavusteinen oppiminen, jossa oppija etsii korpuksista itse tietoa sanaston ja rakenteiden käytöstä, ei ole toistaiseksi rantautunut suomen kielen opetukseen. Lisäksi Akgül (2013: 15) on sitä mieltä, että ”korpuksia tulisi tehdä tunnetuksi suomi

toisena ja vieraana kielenä -opettajien keskuudessa, sillä tällä hetkellä korpuksia hyödynnetään varsin vähän opetuksessa ja opetusmateriaalien laadinnassa”.

2.2. Tutkimusaineisto

2.2.1. Kansainvälinen oppijansuomen korpus

Kansainvälistä oppijansuomen korpusta (*International Corpus of Learner Finnish*, tästä edes ICLFI) on koottu Oulun yliopistossa vuodesta 2007 lähtien, ja se koostuu noin 6 000 tekstistä ja noin miljoonasta saneeista. Siihen on kerätty suomen kielen opiskelijoiden tuottamia tekstejä yli 20 ulkomaisesta yliopistosta, eli aineisto edustaa suomi vieraana kielenä -materiaalia. (Jantunen 2011: 86; Jantunen ym. 2014: 89, 91.)

Keskeisimpien dimensioidensa (ks. Jantunen 2011: 90–92) osalta ICLFI on monitekstiläinen yleiskorpus, joka sisältää sekä fiktiivisiä että ei-fiktiivisiä tekstejä. Siihen on kerätty vain kirjoittamalla tuotettuja tekstejä. Korpus on yksikielinen ja sisältää ainoastaan suomeksi tuotettuja tekstejä. Se ei kuitenkaan sisällä varsinaisia suomennoksia kuten opiskelijoiden tuottamia käännösharjoituksia. ICLFI on pääosin synkroninen korpus, joskin siihen kerätään joistakin opetuspisteistä tekstejä samoilta opiskelijoilta useamman vuosikurssin ajan. Korpus on moniäidinkiäinen, ja sen aineiston tuottajat edustavat kaikkiaan 22:ta eri äidinkieltä. (Jantunen 2011: 90–92; Jantunen & Pirkola 2015: 92–93.) Lisäksi ICLFI sisältää tietoa 22 taustamuuttujasta, jotka koskevat tekstien tuottajia, heidän oppimiskontekstiaan, itse tekstiä sekä muita sekalaisia ominaisuuksia (ks. Jantunen 2011: 93).

ICLFI on monitaitotasoinen korpus. Sen aineisto on jaoteltu taitotasoihin sekä annetun opetustuntimäärän että Eurooppalaisen viitekehyksen (EVK) taitotasoasteikon (A1–C2) mukaan. Tekstien arvioijat ovat olleet kokeneita, koulutettuja EVK:n kriteeristöön perehtyneitä henkilöitä, ja tekstien luokittelu on edellyttänyt vähintään kahden arvioijan samantasoista arviota. On tärkeä huomata, että EVK-perustainen arviointi on teksti- eikä oppijakohtainen, eli saman tekstintuottajan eri tekstit voivat sijoittua eri taitotasoille, ja arviot kertovat yksittäisen tekstin eivätkä oppijan taitotasosta. (Jantunen 2011: 96–97; Jantunen ym. 2014: 67; Brunni, Jantunen & Skantsi 2020: 284.) Toinen huomionarvoinen seikka on, että EVK:n arviointien yhdenmukaisuutta ei ole Jantunen (2011: 97) mukaan

vielä selvitetty, mutta aineistossa on viitteitä siitä, että arviot voivat toisinaan poiketa toisistaan – tosin yleensä vain yhden taitotason verran.

ICLFIN teksteihin on tehty sekä kieliopillinen annotointi että pieneltä osin virheannotointi (ks. Jantunen 2011: 98–100; Jantunen ym. 2014: 68–75; Jantunen & Pirkola 2015: 98–100). Viimeisimpien tietojen (Jantunen & Pirkola 2015: 92–93) mukaan aineiston kieliopillinen annotointi on tehty kokonaan, kun taas virheannotoinnista on tehty 5 %.

Esitän analyysiluvussa edustavan otoksen kokoamistani epätyypillisyyksistä, jotka olen poiminut ICLFIN C-tason raakateksteistä. Käytän tapausten tunnistena yksilöintikoodia, esimerkiksi *PU0022* tai *VI0399a*. Tämä virkkeen lopussa oleva koodi viittaa siihen ICLFIN raakatekstiin, josta kyseinen tapaus on peräisin. Koodin alussa oleva kirjainyhdistelmä viittaa tekstin kirjoittajan äidinkieleen, ja sitä seuraava numerosarja puolestaan yksilöi kirjoittajan. Jos sama kirjoittaja on tuottanut useampia tekstejä, koodin lopussa seuraa kirjain, joka kertoo siitä, monesko teksti tämä on kyseiseltä kirjoittajalta.

2.2.2. Natiivisuomen vertailuaineisto

Koska ICLFI ei sisällä oppijankielen aineistolle verrannollisia natiivituotoksia (Jantunen 2011: 90–91), natiivisuomen vertailuaineisto on löydettävä muualta. Ihanteellinen vertailukohde ICLFille olisi sen taustamuuttujien mukaan muodostettu vertailukorpus, jossa tekstit vastaisivat muodoltaan, sisällöltään sekä keruutavaltaan mahdollisimman pitkälti ICLFIN aineistoa. Sellaista vertailukorpusta ei kuitenkaan ole, joten vertailukohteena toimikoon kirjoitettu yleiskieli. (Brunni ym. 2020: 281.)

Olen valinnut vertailuaineistoksi Ylen suomenkielisen uutisarkiston 2011–2018. Aineisto sijaitsee Kielipankissa, joka on kieliaineistoja käytettävien tutkijoiden palvelukokonaisuus (Kielipankki 2020a). Analysoin aineistoani Kielipankin verkkopohjaisella konkordanssiohjelmalla Korpilla (ks. Borin, Forsberg & Roxendal 2012), jolla pystyn tekemään konkordanssihakuja kieliopillisesti jäsennetystä tekstiaineistosta. On kuitenkin huomiotava, että käyttämäni korpusversio on kirjoitushetkellä vielä beetatestiversio (Kielipankki 2020b).

Ylen suomenkielinen uutisarkisto on jaettu kahdeksaan osakorpukseen vuosien mukaan, ja se sisältää yhteensä 18 708 231 virkettä ja 207 500 221 sanetta. Käytössäni on tutkijoille tarkoitettu versio, jossa virkkeet ovat alkuperäisessä järjestyksessä ja tuki laajennetulle kontekstille on saatavilla. (Yleisradio 2021a.) Huomioitava on myös se, että korpus sisältää tavallisten uutisten lisäksi myös selkouutiset, jotka muodostavat yhteensä 112 485 virkettä ja 826 737 sanetta eli noin 0,6 % koko aineistosta virkemäärältään ja noin 0,4 % sanemäärältään (Yleisradio 2021b). Selkouutisten osuutta ei voida erottaa uutisarkiston muusta aineistosta. En kuitenkaan näe tätä vakavana ongelmana, sillä selkokieltäkin tuotetaan yleiskielen normien puitteissa ja selkouutisten osuus koko korpuksessa on muutenkin pieni.

Olen valinnut vertailuaineistoni useilla eri perusteilla. Ensinnäkin näen, että uutisissa käytetty yleiskieli sopii hyvin vertailukohdaksi ICLFIn teksteille, jotka on myös kirjoitettu pääosin yleiskielellä. Molempien korpusten tekstit ovat myös peräisin suunnilleen samalta ajalta: ICLFIn tekstit vuosilta 2007–2012 ja Ylen suomenkielisen uutisarkiston tekstit vuosilta 2011–2018. Lisäksi yksi ICLFI-aineistoni sisältämä tekstilaji on nimenomaan uutinen, joskin se muodostaa koko oppijansuomen aineistosta vain pienen osan. Vertailuaineiston yhtenä mahdollisena ongelmana on sen yksipuolisuus, sillä ICLFI-aineistoni edustaa uutisten lisäksi myös useita muita tekstilajeja. Siitä huolimatta näen, että kyseinen natiivisuomen aineisto sopii hyvin käyttötarkoituksiini.

2.3. Tutkimusmenetelmät

2.3.1. Epätyypillisyyksien valinta

Aloitan tutkimukseni valitsemalla oppijansuomen aineistostani tarkasteltavia fraseologisia epätyypillisyyksiä. Oppijansuomen aineistoni koostuu kaikista ICLFIn C1- ja C2-taitotasoarvioituista raakateksteistä. Niitä on yhteensä 283, joista 248 on C1-tasoisia ja 35 C2-tasoisia. Olen käynyt kaikki tekstit manuaalisesti läpi vähintään kerran ja poiminut niistä fraseologisiksi epätyypillisyyksiksi mieltämäni tapaukset omaan kieli-intuitiooni nojautuen. Suuren aineistomäärän takia olen rajoittanut käsittelyni niin, että keskityn ainoastaan sananvalintaan, morfologiaan ja ilmaisun rakenteeseen liittyviin tapauksiin. En siis huomioi esimerkiksi tyylillisiä epätyypillisyyksiä, kuten *se*-pronominin yliedustusta

natiivisuomeen verrattuna tai puhekielisyyksiä asiateksteissä. Koska ICLFIn virheannotoinnista on tähän mennessä tehty vain 5 % (Jantunen & Pirkola 2015: 92–93), olen päättänyt sivuuttaa korpuksen virheannotointitiedot kokonaan tulosteni yhdenmukaisuuden säilyttämiseksi.

Koska suomi ei ole äidinkieleni, olen tarkistuttanut alustavat tapaukseni äidinkielisellä kielenkäyttäjällä ja muodostanut analysoitavan epätyypillisyyssokoelmani hänen palautteensa pohjalta. Näin tutkimuksessa on mukana myös natiivin kieli-intuitio oman edistyneen suomenoppijan kieli-intuitioni vastapainona. Samalla varmistan, etten ole ainoa, joka näkee tapauksissa jotain epätyypillistä. On tosiaan huomioitava, että tämä äidinkielinen kielenkäyttäjä on koulutukseltaan suomen kielen ammattilainen, joten hänen kieli-intuitionsa saattaa poiketa kielitieteeseen perehtymättömän äidinkielisen kielenkäyttäjän kieli-intuitiosta.

2.3.2. Epätyypillisyyksien käsittely ja luokittelu

Tutkimuksen seuraavassa vaiheessa käyn valitsemani fraseologiset epätyypillisyydet läpi ja pohdin, miten niitä voi parhaiten kuvailla ja luokitella. Luokittelun lähtökohtana toimivat sananvalinnan, morfologian, lauseketason ja lausetason pääryhmät, ja tarkemmat alaryhmät hahmottuvat vasta tapausten analysoinnin myötä. Odotukseni on, että kunkin ryhmän tapauksista löytyy sekä selkeästi natiivisuomesta poikkeavia tapauksia (ns. prototyyppisiä epätyypillisyyksiä) että tulkinnanvaraisempiakin tapauksia (ns. rajatapauksia), jotka aiheuttavat herkemmin erimielisyyksiä natiivienkin keskuudessa. Koska tapauksia on useita satoja enkä pysty tässä tutkielmassa syventymään jokaiseen epätyypillisyyteen, nostan analyysiosiossani esille vain muutamia kutakin ääripäätä edustavia esimerkkejä kussakin ryhmässä.

Käytän natiivisuomen vertailuaineistoani varmistamaan, että kaikki mainitsemieni tapausten epätyypillisuus on perusteltua ja pitää paikkansa. Esitän kunkin esimerkin rinnalla konkordanssihakutulokseni natiivisuomen vertailukorpuksesta: esimerkiksi epätyypillisen sananvalinnan kohdalla havainnollistan sanan epätyypillistä käyttöä vertailemalla sitä sanan tavanomaiseen esiintymisympäristöön natiivisuomessa. Konkordanssitulosten

avulla pyrin selvittämään alustavia havaintojani empiirisesti sekä saamaan jonkinlaista kuvaa poikkeavuuden asteesta.

Havainnollistan menetelmäni fraseologiseksi epätyypillisyydeksi mieltämälläni tapauksella *tehdä valokuva* (PU0001c) ja sen tyypillisemmällä vastineella *ottaa valokuva*. Teen Korpissa kaksi konkordanssihakua, joista ensimmäinen sisältää lemmat TEHDÄ ja VALOKUVA ja toinen sisältää lemmat OTTAA ja VALOKUVA. Kunkin konkordanssihaun tulokset näyttävät, kuinka monta kerta kyseiset lekseemit esiintyvät toistensa yhteydessä. Vertailemalla tapausten esiintymiskertoja toisiinsa pystyn päättämään, onko suomenoppijan tuottama ilmaus *tehdä valokuva* todellakin niin epätyypillinen natiivisuomessa.

Epätyypillisyyksiä kartoittamalla pyrin löytämään mahdollisia fraseologisia piirteitä, jotka kuvaavat edistynyttä oppijansuomea sekä erottavat sen natiivisuomesta. En tähtää kartoituksessani tyhjentyvyyteen: sen lisäksi, että aihe olisi tämän pro gradu -tutkielman mittakaavassa liian laaja, tyhjentävä kuvaus yhdestäkään oppijansuomen osa-alueesta ei liene edes mahdollista.

2.4. Tutkimukseen liittyvät ongelmat

2.4.1. Kvantitatiivisten menetelmien vähäisyys

Tutkimukseni on pääosin kvalitatiivista, ja yksi mahdollinen ongelma on kvantitatiivisten menetelmien vähäinen rooli analyysissa. Vaikka tukeudunkin isoon määrään autenttista kieliaineistoa kartoittaessani edistyneen oppijansuomen fraseologisia epätyypillisyyksiä ja vertaillen niitä natiivisuomen kielenkäyttöön, analyysini kvantitatiivinen osuus pysyy hyvin yksinkertaisena, enkä perustele tuloksiani varsinaisilla tilastoilla kuten toistuvuus- ja frekvenssilaskelmilla.

Kvantitatiivisen analyysin tärkeyttä on laajasti korostettu korpustutkimuksessa (ks. esim. Tognini-Bonelli 2001), ja harvassa lienevät ne fraseologiset korpustutkimukset, joissa tilastoja ei juurikaan mainita. Itse asiassa muun muassa Biber, Conrad ja Reppen (1998: 8) pitävät kvantitatiivisia menetelmiä olennaisina korpuspohjaisessa tutkimuksessa, ja

yhtenä huolenaiheenani onkin se, uhkaako tilastollisten todisteiden puute tulosteni uskottavuutta.

Tästä huolimatta näen, että tutkimukseni kvantitatiivisen osuuden taso riittää tähtäämääni suuntaa antavaan yleiskartoitukseen, jossa tarkasteltavia tapauksia on paljon ja niitä käsitellään pääpiirteittäin. Tarkoitukseni ei olekaan esittää tilastollisesti merkittävää tietoa fraseologisten epätyypillisyyksien esiintymistiheydestä oppijansuomessa. Sen sijaan pyrin kuvailemaan edistyneen oppijansuomen fraseologisuutta laajasti sekä nostamaan esille joitakin mahdollisesti syvemmän tarkastelun arvoisia piirteitä, jotka esiintyvät kyseisessä kielivariantissa.

2.4.2. Intuition puutteita

Toinen tutkimukseen liittyvä ongelma on se, että nojaudun vahvasti tutkimukseni alkuosassa kieli-intuitioon eli Jantusen (2009a: 101) sanoin rationaaliseen tietoon kielestä sekä sen normeista ja käytöstä. Wasow ja Arnold (2005: 1482) jakavat kielentutkimuksessa käytettävät intuitiot kahteen ryhmään: primaariset intuitiot viittaavat kielenilmauksen muotoiluun tai merkitystä koskeviin päätelmiin, kun taas sekundaarisilla intuitioilla pyritään löytämään selityksiä kyseisiin päätelmiin. Käytän pääsääntöisesti primaarista kieli-intuitiota valitessani analysoitavia fraseologisia epätyypillisyyksiä ICLFIn suomenoppijoiden teksteistä.

Vaikka kielentutkimuksessa on jouduttu turvautumaan pitkälti intuitioon ja siihen pohjautuvaan introspektioon ennen korpusten tuloa, ne eivät kerro todellisesta kielenkäytöstä vaan pikemminkin siitä, mitä kielenkäyttäjät tietävät kielestä ja millaiseksi he mieltävät kieltä (Tsui 2004: 39). Näin intuition ja introspektion yksinomainen käyttäminen tutkimusmenetelmänä on useiden tutkijoiden (ks. esim. Schütze 2016; Gries 2002; Wasow & Arnold 2005) mukaan osoittautunut puutteelliseksi, ja pelkästään niiden avulla on vaikea saada luotettavaa tietoa kielenkäytöstä (Sinclair 2004a: 2; Mauranen 2004: 95–96) tai kuvata kieltä riittävän tarkasti (Ojutkangas 2008: 277). Myös Biber, Conrad ja Reppen (1998: 3, 9) ovat sitä mieltä, että kielenkäyttötapoja tutkittaessa intuition tai yksittäistapauksen varassa tehdyt johtopäätökset voivat olla epäluotettavia, sillä ihmisillä on tapana huomata poikkeavia tapauksia enemmän kuin tyypillisiä. Lisäksi Tsui (2004: 40) ja

Francis (1993: 139) huomauttavat, että monia kielenkäyttöä koskevia ilmiöitä ei edes pystytä havaitsemaan introspektiolla.

Puutteistaan huolimatta intuitiolla ja introspektiolla on paikkansa kielentutkimuksessa. Esimerkiksi Chafe (1992: 86) huomauttaa, että kielentutkimuksessa on nojaututtu ja nojaututaan edelleenkin paljolti introspektioon. Biber, Conrad ja Reppen (1998: 10) toteavat, että intuitio ja yksittäistapauksiin pohjautuvat havainnot voivat johtaa mielenkiintoiisiin korpuspohjaisiin tutkimuksiin. Muun muassa Stubbs (1995: 48; 2002: 9) ja Jantunen (2009a: 107–109) pitävät tutkijan intuitiota jopa välttämättömänä osana empiiristä tutkimusta, sillä se ohjaa pitkälti tutkimuskysymysten muodostamista, tutkimusten toteuttamista sekä tutkimustulosten tulkintaa. Samoilla linjoilla on Cruse (1986: 10–12), jonka mukaan se, kuulostaako tietty ilmaus oudolta äidinkielen kielenkäyttäjän korvissa, on yksinkertaisimpia semanttisia päätelmiä, joihin kielenkäyttäjä kykenee. Lisäksi Gries (2002: 27) toteaa, että intuitio voi toimia hyvänä, käteväenä ja jopa tarkkana analyysin lähtökohtana, josta syntyy arvokkaita havaintoja kieli-ilmiöstä.

On siis selvää, että intuitiosta ja introspektiosta voi olla hyötyä kielentutkimuksessa, kunhan niiden avulla tehtyjen havaintojen paikkansapitävyyttä selvitetään myös empiirisin menetelmin. Näin ovatkin monet tutkijat todenneet: esimerkiksi sekä Sinclair (2004b: 279) että Wasow ja Arnold (2005: 1485) pitävät intuitiota käypänä välineenä hypoteesien muodostamisessa, jonka jälkeen sitä täydennetään muilla tieteellisillä menetelmillä. Myös Gries (2002: 27) korostaa, että intuition lisäksi tutkijan on hyödynnettävä metodologisesti luotettavampia tutkimusmenetelmiä sekä oltava valmis päivittämään hypoteesejaan korpusaineistosta nousevan tiedon mukaan.

Korpuukset ovat viime vuosikymmeninä nousseet tärkeäksi intuitiota täydentäväksi – ja joskus sen kumoavaksi – tekijäksi kielentutkimuksessa. Muun muassa Francis (1993: 139) painottaa, että pyrittäessä kuvaamaan kieltä painoarvo on annettava ensisijaisesti korpusaineistolle eikä intuitiolle, sillä usein on ero sen välillä, mitä ihmiset ajattelevat sanovansa ja mitä he oikeasti sanovat. Samoilla linjoilla ovat Biber, Conrad ja Reppen (1998: 3, 9), joiden mukaan kielentutkimuksessa on analysoitava runsasta, lukuisten kielenkäyttäjien tuottamaa autenttista aineistoa, jottei johtopäätöksiä tehtäisi pienen joukon ominaispiirteiden perusteella.

Tiivistäen voidaan sanoa, että introspektio- ja aineistometodi ovat lopulta molemmat tarpeellisia, toisiaan täydentäviä lähestymistapoja kielentutkimuksessa (Biber, Conrad & Reppen 1998: 10; Granger & Meunier 2008: xxvii; Jantunen 2009a: 107). On siis turhaa väitellä siitä, kumpi on kielentutkimuksessa parempi tai yrittää korvata yhtä toisella.

Oman tutkimukseni kannalta näen Kuuluvaisen (2015: 2, 23) tavoin, että kieli-intuition käyttö fraseologisten epätyypillisyyksien käsittelyssä on haasteen lisäksi myös mahdollisuus, jopa välttämättömyys. Kuuluvainen (mts. 23) mainitsee, että ICLF:n virheannotoinnissa ja -luokituksessa fraseologisiksi virheiksi ovat päättyneet ”kaikenlaiset natiivikielille epätyypilliset, usein vierasvaikutteiset ilmaiset”, joiden luokittelun perustaksi on riittänyt annotoijan subjektiivinen, kieli-intuition perustuva tuntemus ilmaisun oudon kuulaisuudesta. Näen siis kieli-intuition käyttämistä tarpeellisenä varsinkin tutkimukseni alkuvaiheessa, epätyypillisyyksien tunnistamisessa.

Jotta tutkimustulokseni eivät olisi liikaa primaarisen kieli-intuition varassa ja siten mahdollisesti heikosti perusteltavissa, olen ottanut käyttööni natiivisuomen vertailuaineiston, jonka avulla voin joko vahvistaa tai kumota hypoteesini valitsemieni tapausten fraseologisesta epätyypillisyydestä. Noudatan tutkimuksessani samaa linjaa kuin esimerkiksi Wasow ja Arnold (2005: 1488), eli pidän intuitiota tärkeänä tekijänä alustavien havaintojen tekemisessä, mutta annan empiiriselle todistusaineistolle enemmän painoarvoa. En siis näe intuition käytössäni suurta ongelmaa niin kauan kuin pidän Francisin (1993: 139) tavoin korpusta ainoana luotettavana tiedonlähteenä ja olen valmis muokkaamaan lähtöoletuksiani uuden korpustodisteen valossa.

3. TEORIA JA TAUSTA

3.1. Kontekstuaalinen semantiikka

3.1.1. Käsitteen tausta ja määritelmä

Semantiikka eli merkitysoppi on kielentarkastelun osa, jossa tutkitaan ilmausten merkityksiä. Sen piiriin kuuluu kaksi tutkimuksenikin kannalta keskeistä osa-aluetta: **sanase-mantiikka** eli **leksikaalinen semantiikka** tutkii sanojen merkityksiä ja merkityssuhteita, kun taas **lausesemantiikka** eli **syntaktinen semantiikka** keskittyy lauseiden merkitysten ja merkityssuhteiden tarkasteluun. Osa-alueet eivät kuitenkaan ole kokonaan erotettavissa toisistaan, sillä lauseen merkitys muodostuu lauseessa esiintyvien sanojen pohjalta, ja sanan täsmällinen merkitys puolestaan selviää vasta kun se nähdään itseään laajemmassa kontekstissa. (Häkkinen 1994: 165–167; Kangasniemi 1997: 22.)

Sana- ja lausesemantiikan lisäksi tutkimukseni teoreettiseen taustaan kuuluu **kontekstuaalinen semantiikka**, jossa merkityksen muodostumista tarkastellaan kontekstilähtöisesti. Kontekstin laajuus vaihtelee tutkimuksittain: se voi olla suppeimmillaan ilmausta ympäröivä teksti tai laajimmillaan tarkastelukohteen koko sosiaalinen ympäristö. Kontekstuaalisessa semantiikassa korostetaan merkityksen ja muodon sekä leksikon ja kielipin välistä yhteyttä, ja tutkimuksen keskipisteessä ovat sanojen konkreettiset, toistuvat käyttöyhteydet. Kyse on kuitenkin paljon laajemmasta ilmiöstä kuin vain sanojen yhteisi- esiintymisestä, sillä tarkasteltavat suhteet eivät ole pelkästään leksikaalisia, vaan niissä on myös kielipillinen ja semanttinen ulottuvuus. (Jantunen 2004: 7–9.)

Kontekstuaalisen semantiikan lähestymistapa juontaa juurensa niin kutsutusta brittiläisestä koulukunnasta, jonka merkittävimmät edustajat ovat J. R. Firth, M. A. K. Halliday sekä John Sinclair. Koulukunnan peruseriaatteisiin kuuluu muun muassa se, että kieltä tulee tarkastella todellisten kielenkäyttöesimerkkien pohjalta, kokonaisina teksteinä eikä kontekstistaan erotettuina katkelmina. Lisäksi leksikkoa ja kielioppia pidetään toisistaan riippuvaisina, eikä muodon ja merkityksen nähdä olevan erotettavissa toisistaan. (Stubbs 1996: 22–23.)

3.1.2. Tilannekonteksti, konteksti ja koteksti

Keskeinen käsite Firthin (1957, 1986 [1957]) kieliteorialle on **tilannekonteksti** (*context of situation*). Itse käsite on abstrakti, mutta se on kytköksissä todellisiin, autenttisiin tapahtumiin ja kokemuksiin. Tilannekontekstin voidaan nähdä muodostuvan erilaisista kategorioista, joista kielentutkimuksen kannalta olennaiset ovat muun muassa tilanteessa toimivat henkilöt sekä heidän verbaalinen ja ei-verbaalinen viestintänsä. Tarkastelun kohteena ovat yksittäisten vuorovaikutustapahtumien sijaan tyypilliset, rutiininomaiset sekä toistuvasti havaittavat kielenkäyttötilanteet. Jokainen teksti on siis osa sitä laajempaa tilannekontekstia, ja jokainen ilmaus saa lopullisen merkityksensä niin tekstinsisäisistä kuin -ulkoisistakin tekijöistä. (Firth 1957: 35, 182; 1968 [1957]: 175–177.)

Firthin käsitys tilannekontekstista perustuu siihen, että kieli ja ihmisen muut sosiaaliset toiminnot ovat toisistaan erottamattomia (Jantunen 2004: 10). **Konteksti** (*context*) taas viittaa tilannekontekstia suppeampaan kokonaisuuteen, joskin termi on monitulkintainen eikä raja kahden käsitteen välillä ei ole aina selkeä. Konteksti voi vaihdella tarkasteltavan kohteen konkreettisesta tekstiympäristöstä laajempaan sosiokulttuuriseen taustaan, jonka tekijät ovat mukana vaikuttamassa tekstin tulkintaan. (Sinclair 1991: 171; Jantunen 2004: 11–12.) Lisäksi kontekstin piirissä olevien tekijöiden nimeäminen voi olla hankalaa, sillä aina ei ole selvää, milloin teksti loppuu ja mistä konteksti alkaa (Heikkinen 1999: 68).

Siinä missä tilannekontekstin tarkastelun piiriin kuuluvat kielenulkoiset tekijät ja kontekstin voidaan katsoa huomioivan myös tekstinulkoisia tekijöitä merkityksen tulkinnassa, **koteksti** (*co-text*) viittaa puolestaan tekstinsisäisiin tekijöihin, sanan tai ilmauksen molemminpuoliseen tekstuaaliseen lähiympäristöön (Sinclair 1991: 172). Voidaan siis sanoa, että koteksti on näistä kolmesta käsitteestä suppein ja tarkkarajaisin. Jantunen (2004: 12) lisää kotekstin määritelmään sen, että siihen voi kuulua lauseyhteyden lisäksi myös lause- ja virkerajat ylittävä teksti. Siksi korpuslingvististä analyysia tehtäessä olisi hänen mukaansa syytä määritellä tarkastelukoteksti eli se tekstialue, jonka avulla ilmausten koteksuaalisia piirteitä tarkastellaan.

3.1.3. Syntagmaattiset ja paradigmaattiset suhteet

Kontekstuaalisen semantiikan käsitteeseen kuuluu ajatus, että kielellinen yksikkö – tässä tapauksessa sana tai ilmaus – on kahdenlaisessa suhteessa muihin samalla kielellisellä tasolla oleviin yksikköihin. Yhtäältä sana on **syntagmaattisessa suhteessa** muihin sanoihin, joiden kanssa se muodostaa lineaarisen jatkumon eli lauseen tai virkkeen. Toisaalta sana on **paradigmaattisessa suhteessa** niihin muihin sanoihin, joiden kanssa se on vaihdettavissa saman lauseenjäsenen roolissa. (Lyons 1968: 73–74; Cruse 2000: 148–149; Jantunen 2004: 8; Karlsson 2008: 18.)

Suhteita havainnollistaa ICLFIstä poimittu ja hieman muokattu lausekatkelma <monet Kaarlen vastustajista joutuivat linnaan> (PU0022a). Lausetta tarkasteltaessa voidaan sanoa, että <MONI>, <KAARLE>, <VASTUSTAJA>, <JOUTUA> ja <LINNA> ovat syntagmaattisessa suhteessa keskenään, sillä ne muodostavat yhdessä kieliopillisesti ja semanttisesti ymmärrettävän lauseen. <LINNA> voi puolestaan olla paradigmaattisessa suhteessa sellaisten sanojen kuin <VANKILA> tai <TYRMÄ> kanssa, sillä ne kaikki viittaavat suunnilleen samantyyppiseen tarkoitteeseen.

Syntagmaattisia ja paradigmaattisia suhteita ohjaa **kontekstuaalinen valinta** eli Jantusen (2004: 8) sanoin ne säännöt tai preferenssit, joiden perusteella jokin sana valitaan paradigmaattisista suhteistaan syntagmaattiseen suhteeseen. Kyseessä on jatkumo, jonka toisessa ääripäässä vallitsee sanojen täysin vapaa valinta, toisessa taas niiden täysin rajoittunut myötäesiintyminen. Ääripäiden välissä on niitä sanojen myötäesiintymiä, joita ohjaavat rajoituksia lievemmat preferenssit. Ero valintarajoitusten ja -preferenssien välillä on se, että edellinen viittaa sanojen *mahdollisiin* myötäesiintymiin ja perustuu joko–tai-suhteeseen, kun taas jälkimmäinen viittaa sanojen *tyypillisiin* myötäesiintymiin ja sallii jossain määrin variaatiota. (Jantunen 2004: 15.)

3.1.4. Vapaan valinnan periaate ja idiomiperiaate

Sanojen kontekstuaaliseen valintaan liittyy myös Sinclairin (1991) esittämä kaksi yhdistymisperiaatetta, joilla voidaan selittää merkityksen syntymistä tekstistä. **Vapaan valinnan periaate** (*open-choice principle*) on näkemys, jonka mukaan teksti syntyy lukuisten

monimutkaisten valintojen tuloksena ja kieliopillisuus on ainoa rajoittava tekijä valintoja tehtäessä. Teksti nähdään aukkosarjana, jossa aukot täytetään jollain sopivan leksikon ehdokkaalla, ja vaihtoehdot ovat miltei rajattomat. On kuitenkin selvää, että sanat eivät päädy tekstiin satunnaisin perustein, eivätkä vapaan valinnan mallin rajoitukset riitä selittämään tekstin muodostumista. (Sinclair 1991: 109–110.)

Sinclair nostaa toki esille joitakin valinnanvaraa rajoittavia tekijöitä, jotka toimivat vapaan valinnan periaatteen puitteissa. Hän toteaa, että joillakin kielenulkoisessa maailmassa yhdessä esiintyvillä asioilla on taipumus valikoitua samaan joukkoon kielessäkin, joskin kieli sisältää myös runsaasti valintoja, joihin kielenulkoisen todellisuus ei vaikuta. Lisäksi hän mainitsee rekisterit eli tilanteenmukaiset kielenkäyttötavat, jotka sisältävät kukin omia pienempiä kielellisiä valintajoukkoja. (Sinclair 1991: 110.)

Näistä valinnanvaraa supistavista tekijöistä huolimatta Sinclair kuitenkin näkee, että vapaan valinnan periaate sisältää liian paljon valinnanmahdollisuuksia ollakseen toimiva selitys sanojen yhdistymiselle. Ongelman ratkaisuksi hän tarjoaa vaihtoehtoisen mallin, jolla ilmiötä voidaan selittää aukottomammin. **Idiomiperiaate** (*idiom principle*) viittaa näkemykseen, että kielenkäyttäjällä on käytössään iso määrä puolivalmiita konstruktioita (*semi-preconstructed phrases*), joista kukin edustaa kokonaisuudessaan yhtä valintayksikköä. Yksinkertaisimmillaan idiomiperiaate ilmenee kaksisanaaisena ilmaisuna, joka toimii useampiosaisuudestaan huolimatta yhtenä kiinteänä merkitysyksikkönä. Idiomiperiaatteen mukaan juuri sellaisten konstruktoiden katsotaan ohjaavan yksittäisten sanojen käyttöä ja tekstin rakentumista. (Sinclair 1991: 110–111.)

Sinclair painottaa, että vapaan valinnan periaatteen ja idiomiperiaatteen välillä on selkeä raja, eivätkä mallit ole yhteensovittavissa toistensa kanssa. Hän katsoo, että tavalliset tekstit rakentuvat ensisijaisesti idiomiperiaatteen mukaan ja tekstiä tulkittaessa siirrytään vapaan valinnan periaatteen varaan vain silloin, kun vastaan tulee esimerkiksi odotustenvastainen sananvalinta. Sinclair pitää siis vapaan valinnan periaatetta toissijaisena mallina, joskin siitä ei myöskään voi hänen mielestään luopua kokonaan. (Sinclair 1991: 114.) Itse odotan, että vapaan valinnan periaatteen vaikutus tulee näkymään selkeästi aineistossani, sillä epätyypillisyyksiä syntyy nimenomaan silloin, kun suomenoppija ei kykene muodostamaan kohdekielensä ilmauksia idiomiperiaatteen mukaan natiivin tavoin.

3.1.5. Leksikaalinen ja morfologinen priming

Michael Hoeyn (2005: 7–8) mukaan sanojen toistuvien yhteisesiintymisten taustalla on psykologinen ilmiö, jonka kuvaamiseksi hän lainaa psykologian käsitettä **priming**. Hän katsoo, että joka kerta kun kielenkäyttäjä kohtaa vuorovaikutuksessa tietyn sanan, hänen mieleensä tallentuu sanan lisäksi myös sen silloinen käyttöympäristö. Näin sanan kumulatiiviset esiintymät ja käyttökontekstit latautuvat sen merkitykseen, ja se ikään kuin virittäytyy esiintymään kielenkäyttäjän mielessä tiettyjen sanojen kanssa tietyissä konteksteissa. Tähän ilmiöön Hoey viittaa termillä **leksikaalinen priming**.

Tässä kohdassa on huomautettava, että Hoeyn priming-käsite on pidettävä erillään psykologian ja psykolingvistiikan priming-käsitteistä, jotka molemmat koskevat ärsykkeeseen perustuvaa reagointia ja virittäytymistä (Jantunen & Brunni 2012: 73–74). Kun käytän *priming*-termiä tässä tutkielmassa, viittaan ainoastaan Hoeyn kuvaamaan ilmiöön.

Hoeyn mukaan sanan primingia ei tulisi nähdä muuttumattomana, vaan sanan joka esiintymis- ja käyttökerta joko vahvistaa tai heikentää sitä. Jos sana kohdataan tai sitä käytetään entuudestaan tutussa käyttöympäristössä, kokemus vahvistaa sanan assosiaatiota käyttöympäristöönsä ja siten myös sen primingia. Päinvastoin jos sana kohdataan tai sitä käytetään ennestään vieraassa käyttöympäristössä, kokemus heikentää sanan assosiaatiota siihenastiseen käyttöympäristöönsä ja siten myös sen primingia. (Hoey 2005: 9.)

Lisäksi primingissa voi ilmetä säröjä (*cracks*), jotka aiheutuvat esimerkiksi silloin kun sanaan on karttunut tietty priming kielenkäyttäjän mielessä, mutta joku toteaa sen virheelliseksi. Näitä säröjä voidaan paikata joko kumoamalla sanan alkuperäinen priming tai torjumalla siihen kohdistuva korjaus. Säröjen olemassaolon mahdollisuus antaa ymmärtää, että jokaisen kielenkäyttäjän kokemukset ja primingit ovat ainutlaatuisia, joskin on myös oltava sääntöjä tai preferenssejä, jotka yhtenäistävät yksittäisten kielenkäyttäjien primingit vuorovaikutuksen sujuvoittamiseksi. (Hoey 2005: 11.)

Muun muassa Jantunen ja Brunni (2012: 74–75) ovat esittäneet, että priming liittyy laajasti fraseologiaan (ks. luku 3.2.). Heidän mukaansa ilmausten odotuksenmukaiset käyttöyhteydet vaihtelevat tyypillisistä kollokaateista (ks. luku 3.2.3.1.) tyyllillisiin ja tekstuaalisiin preferensseihin, ja näiden käyttöyhteyksien omaksumisessa juuri toistuvuudella

on suuri merkitys. Niinpä priming näyttäytyy natiivin kielenkäytössä toistuvina käyttöyhteyksinä, sillä äidinkielen kielenkäyttäjä on hyvin perillä siitä sanan käyttöä koskevasta tiedosta, joka kuuluu sanan fraseologisuuteen. (Mp.)

Jantunen ja Brunni (2012) ovat myös tutkineet priming-käsitteen morfologista ulottuvuutta. **Morfologinen priming** koskee heidän mukaansa ”taivutusmuotojen käyttöä ja muotoihin liittyviä kontekstuaalisia piirteitä”, ja heidän pilottiluonteisen tutkimuksensa tulokset paljastavat morfologisen primingin ja fraseologian välisen syy–seuraussuhteen. Fraseologisen yksikön morfologinen priming tarkoittaa sitä, että sekä fraseologisen yksikön ytimellä (*core*) että sen kontekstuaalisilla kerailmauksilla on tietynlainen odotuksenmukainen priming johonkin taivutusmuotoon. Esimerkiksi PITÄÄ-verbiä tutkiessaan Jantunen ja Brunni ovat havainneet, että tuotettujen muotojen kirjo on kapeampi suomenoppijoilla kuin natiiveilla ja oppijansuomessa yleisiä *pidän-* ja *pitää-*muotoja esiintyy huomattavasti harvemmin natiivisuomessa. Lisäksi PITÄÄ-verbin käyttö merkityksessä ’tykätä’ on yliedustunut oppijansuomessa, kun taas sen käyttö arviointimuotissa ’pitää jotakin jonakin’ on aliedustunut natiivisuomeen verrattuna. (Jantunen & Brunni 2012: 71, 77–78, 88–90, 94.)

Jantunen ja Brunni (2012: 73, 95) pitävät morfofonologisten piirteiden tarkastelua olennaisena oppijankielen fraseologisessa tutkimuksessa, varsinkin kun kyseessä on taivutukseltaan monipuolinen kieli, kuten suomi. Fraseologian ja morfologisen primingin välinen yhteys tarkoittaa sitä, että kielenoppijan priming eroaa natiivin primingista niin leksikaalisesti kuin morfologisestikin (mts. 94). Koska suomenoppijoiden tuottamiin fraseologisiin epätyypillisyyksiin voi sisältyä myös morfologisesti poikkeavia tapauksia, kiinnitän tutkimuksessani huomiota sanan lemmatason eli perusmuodon lisäksi myös sen muihin taivutusmuotoihin.

3.2. Fraseologia ja fraseologinen yksikkö

3.2.1. Käsitteen tausta ja määritelmä

Vaikka monisanaisia yksiköitä eli sanayhdistelmiä (*multi-word units*) on tutkittu jo kauan, fraseologia alkoi vakiintua omaksi tutkimusalakseen vasta viime vuosikymmenien aikana.

Sen myötä kiinnostus fraseologiaa kohtaan on kasvanut kovalla vauhdilla entisestään. Fraseologian vakiintumista ovat kuitenkin hankaloittaneet alan laajamittaisuus ja sillä vallitseva vaihtelevuus sekä sen sekava ja paikoittain ristiriitainen terminologia. (Granger & Meunier 2008: xix–xx; Granger & Paquot 2008: 27–28; Gries 2008: 3.)

Granger ja Paquot (2008: 28–29) erottavat kaksi päähaaraa fraseologisen tutkimuksen alalla. Ensimmäinen, perinteinen päähaara, jota Nesselhauf (2004a) kutsuu **fraseologiseksi lähestymistavaksi** (*phraseological approach*), juontaa juurensa Itä-Euroopasta. Lähestymistavassa fraseologista yksikköä on pyritty määrittelemään ennalta sovittujen kriteerien perusteella, ja siinä on usein keskitytty kiinteisiin sanayhdistelmiin, kuten idiomeihin ja sanontoihin. Fraseologia on nähty jatkumona, jonka yhdessä ääripäässä ovat jähmeimmät ja läpinäkyvämmät tapaukset ja toisessa ääripäässä muuttuvimmat ja läpinäkyvimmat tapaukset. Fraseologian ’ytimenä’ on pidetty niitä idiomaattisimpia tapauksia, joiden merkitykset eivät aukea niiden osien summasta. (Granger & Paquot mp.)

Toinen, uudempi päähaara, jota Nesselhauf (2004a) kutsuu **frekvenssipohjaiseksi lähestymistavaksi** (*frequency-based approach*), on peräisin brittiläisestä koulukunnasta, erityisesti John Sinclairin työstä leksikografian parissa. Sen sijaan että lähtökohtana ovat ennalta määrätyt kriteerit, lähestymistavassa pyritään löytämään leksikaalisia myötäesiintymiä induktiivisesti korpusaineistojen avulla. Fraseologian määritelmä on laajennettu sisältämään myös monia sanayhdistelmiä, jotka eivät perinteisesti kuuluneet fraseologian piiriin. Uuden lähestymistavan valossa monet aiemmin rajatapauksina pidetyt yksiköt ovat nousseet keskeiseen asemaan esiintymistiheytensä perusteella, kun taas monet perinteisesti prototyyppiset yksiköt ovat osoittautuneet hyvin harvinaisiksi ja siten tutkimuksen kannalta toisarvoisiksi. (Granger & Paquot 2008: 28–29.)

Lähestymistapojen moninaisuus on johtanut vaikeuksiin fraseologian tarkassa määrittelyssä. Grangerin ja Meunierin (2008: xix–xx) mukaan fraseologiasta puhutaan yleensä leksikologian osa-alueena, jossa tutkitaan yksittäisten sanojen sijaan sanayhdistelmiä. Siinä missä tutkimuksen keskiössä olivat aiemmin vain kaikista jähmeimmät ja läpinäkyvämmät tapaukset, nykyään fraseologia on laajentunut kattamaan myös ne leksikaaliset yksiköt, jotka olisivat perinteisesti jääneet fraseologian piirin ulkopuolelle (mp.). Jantunen (2009b: 360) mainitseekin, että fraseologisen näkökulman lähtökohtana on ajatus siitä, että iso osa kielenkäytöstä on nimenomaan fraseologista.

Noudatan tässäkin tutkimuksessa fraseologian nykyistä, laajempaa määritelmää. Käsitän siis fraseologisuuden Jantusen ja Brunnin (2012: 75) tavoin tekstuaalisena ilmiönä sanastollisine, kieliopillisine ja semanttisine ulottuvuuksineen. Grangerin (mm. 1998a) ja Jantusen (2009b: 375, 360–361) tavoin sisällytän fraseologian piiriin fraasien, sanontojen, idiomien ja keskustelurutiinien lisäksi myös muut vakiintuneet monisanaiset leksikaaliset yksiköt ja viittaan kaikkiin näihin tapauksiin termillä **fraseologinen yksikkö** (tarkkarajaisemmasta määritelmästä ks. seuraava luku).

3.2.2. Fraseologinen yksikkö

Ytimekkäimmillään fraseologinen yksikkö on Jantusen (2009b: 358) sanoin ”yhtä sanaa laajempi elementti”. Määritelläkseni sen tarkemmin käytän Griesin (2008: 4–10) esittelemää kuutta parametria, jotka koskevat fraseologisen yksikön ominaisuuksia. Niiden mukaan se, lasketaanko sanojen myötäesiintymä fraseologiseksi yksiköksi, voidaan määritellä seuraavilla kriteereillä:

1. myötäesiintymän osien luonne,
2. myötäesiintymän osien määrä,
3. myötäesiintymän tarvittava esiintymismäärä ennen kuin sen voi laskea fraseologiseksi yksiköksi,
4. myötäesiintymän osien etäisyys toisistaan,
5. myötäesiintymän osien leksikaalinen ja syntaktinen joustavuus (*lexical and syntactic flexibility*) sekä
6. myötäesiintymän semanttinen yhtenäisyys (*semantic unity*) ja semanttinen epäkompositionaalisuus (*semantic non-compositionality*) (Gries mts. 4).

Myötäesiintymän osien luonne. Gries (2008: 5) laskee fraseologisiksi yksiköiksi kaikki myötäesiintymät, jotka koostuvat yhdestä leksikaalisen yksikön lemmasta tai sen muodosta sekä mistä tahansa toisesta kielellisestä elementistä. Hän (mts. 5–6) korostaa, että fraseologisessa yksikössä on oltava nimenomaan vähintään yksi leksikaalisesti spesifioitu (*lexically specified*) elementti eli sellainen elementti, jonka tarkka lekseemi on tiedossa (vrt. yleiset elementit kuten sanaluokat ja lauseenjäsenet). Myötäesiintymän toinen osa voi taas olla esimerkiksi toinen leksikaalinen yksikkö tai jokin kieliopillinen rakenne

(*grammatical pattern*, vrt. *grammatical relation*). Samoilla linjoilla Griesin kanssa on Jantunen (2012a: 205), joka kutsuu fraseologista yksikköä lekseemin ja sen ko(n)tekstuaaliset valinnat sisältäväksi leksikaalis-kieliopilliseksi rakenteeksi. Sekä Gries että Jantunen käsittävät fraseologisen yksikön huomattavasti laajempänä ilmiönä kuin mihin sillä on perinteisesti viitattu (esim. idiomeihin ja kiinteisiin fraaseihin), ja otan tässäkin tutkimuksessa heidän parametrinsa käyttöön sellaisenaan.

Myötäesiintymän osien määrä. Toinen parametri koskee sitä, voiko fraseologinen yksikkö koostua vain kahdesta elementistä vai voiko siihen sisältyä useampikin elementti. Kaksi- ja useampisanaiset myötäesiintymät onkin Jantunen (2004: 18) mukaan usein erotettu toisistaan tarkastelussa, ja niille on myös annettu eri nimityksiä (ks. esim. Jantunen mts. 19; 2009b: 359). Yhdyn tässä tutkimuksessa Griesin (2008: 5) näkemykseen siitä, että fraseologinen yksikkö voi koostua useammasta kuin kahdesta osasta.

Myötäesiintymän tarvittava havaitsemismäärä. Gries (2008: 5) laskee fraseologisiksi yksiköiksi tapaukset, joiden havaittu esiintymisfrekvenssi on suurempi kuin niiden odotettu esiintymisfrekvenssi. Kyseinen määritelmä antaa ymmärtää, että myötäesiintymän fraseologisuuden arvioimiseksi on ensin nähtävä tietoja sen suhteellisista esiintymisfrekvensseistä. Kriteeri on tämän tutkimuksen kannalta ongelmallinen siinä mielessä, ettei minulla ole käytössäni tilastollisia testejä, joihin voin perustaa arviointini. Joudun siis muuntamaan parametria niin, että jätän suhteelliset frekvenssilaskelmat pois kriteereistä ja perustan arviointini sellaiseen vakiintuneisuuteen, joka on pääteltävissä myötäesiintymien absoluuttisista esiintymiskerroista. Stubbs (1995: 29) mainitseekin, että kvantitatiivista lingvististä tietoa käsiteltäessä tapausten yksinkertainen yhteen laskeminen ja luetelu usein riittänee. Parametrini mukaan myötäesiintymä on siis riittävän vakiintunut ja siten laskettavissa fraseologiseksi yksiköksi, jos korpuksen konkordanssihaun kvalitatiivisesti tarkistetut tulokset viittaavat yksittäistapausta yleisempään käyttöön.

Myötäesiintymän osien sallittu etäisyys toisistaan. Gries (2008: 5) mainitsee, että joissain tutkimuksissa fraseologisiksi yksiköiksi lasketaan vain sellaiset myötäesiintymät, joiden osat esiintyvät vierekkäin. Hän kuitenkin sisällyttää määritelmäänsä nekin tapaukset, joiden osat esiintyvät kauempana toisistaan. Puhuessaan kollokaateista (ks. luku 3.2.3.1.) Stubbs (1995: 32–33) toteaa, ettei tutkimuskirjallisuudessa ole toistaiseksi päästy yhteisymmärrykseen sopivasta myötäesiintymän havaintoalueesta. Hänen

mukaansa yleisesti käytetyt havaintoalueet ovat kaksi tai kolme sanaa hakusanan molemmin puolin (2:2 tai 3:3), kun taas esimerkiksi Sinclair (1991) pitää optimaalisena havaintoalueena korkeintaan neljää sanaa hakusanan molemmin puolin (4:4). Lisäksi Halliday (1966: 151) huomauttaa, että kollokaatiosuhteet voivat joskus ylittää virkerajat. Pitäytyäkseni pro gradu -tutkielman laajuudessa otan käyttöön fraseologisen yksikön havaintoalueeksi virkkeen laajuisen kontekstin ja tarkastelen vain niitä myötäesiintymiä, jotka esiintyvät samassa virkkeessä. Griesin (mp.) tapaan en edellytä määritelmässäni myötäesiintymien osien vierekkäisyyttä.

Myötäesiintymän osien leksikaalinen ja syntaktinen joustavuus. Tässä parametrissa päätetään siitä, missä määrin myötäesiintymien osissa sallitaan leksikaalista ja syntaktista vaihtelua. Täysin joustamattomia tapauksia ovat sellaiset myötäesiintymät, joissa ei esiinny minkäänlaista vaihtelua. Esimerkiksi ilmaus *yhteen ääneen* esiintyy aina sellaiseenaan yksikön illatiivimuodossa. Täysin joustavia tapauksia ovat taas sellaiset rakenteet, jotka eivät sisällä yhtäkään leksikaalisesti spesifioitua elementtiä ja jonka kaikki osat ovat vapaasti vaihdettavissa rakenteen sääntöjen puitteissa. Tällainen rakenne on esimerkiksi transitiivinen muotti [SUB VER OBJ], johon sopii periaatteessa mikä tahansa subjekti, transitiiviverbi ja objekti. Jossain määrin joustavia tapauksia ovat puolestaan sellaiset myötäesiintymät, joiden yhdessä tai useammassa osassa esiintyy jotain vaihtelua. Esimerkiksi ilmauksessa *potkaista tyhjää* verbi *potkaista* voi esiintyä eri aika- ja persoonamuodoissa mutta *tyhjää* pysyy aina partitiivimuodossa. Kuten Grieskin (2008: 5–6), sallin omassa määritelmässäni leksikaalista ja syntaktista joustavuutta myötäesiintymien osissa ja laskeen ne fraseologisiksi yksiköiksi, kunhan ne sisältävät vähintään yhden leksikaalisesti spesifioitun elementin.

Myötäesiintymän semanttinen yhtenäisyys ja semanttinen epäkompositionaalisuus. Semanttinen yhtenäisyys tarkoittaa sitä, että useampiosaisuudestaan huolimatta myötäesiintymä toimii yhtenä semanttisena yksikkönä, eli se välittää merkitystä ikään kuin se olisi vain yksi sana (Gries 2008: 6). Tällaisia ovat esimerkiksi ilmaukset *suoraa päätä* ja *tieten tahtoen*. Semanttinen epäkompositionaalisuus tarkoittaa puolestaan sitä, että myötäesiintymän merkitys ei ole suoraan pääteltävissä sen osien summasta (Granger & Paquot 2008: 31). Tämä pätee esimerkiksi prototyyppeihin idiomeihin (ks. esim. Nenonen 2002), kuten *pää kolmantena jalkana*. Omassa fraseologisen yksikön määritelmässään Gries (mp.) edellyttää myötäesiintymissä semanttista yhtenäisyyttä muttei semanttista

epäkompositionaalisuutta. Käytän omassa tutkimuksessani samaa parametria kuin Gries, sillä semanttisen epäkompositionaalisuuden vaatiminen sulkisi pois monia tapauksia, jotka täyttävät muuten fraseologisuuden laajemman määritelmän kriteerit.

Yhteenvedona määrittelen tässä tutkimuksessa fraseologisen yksikön seuraavasti: se on sanojen myötäesiintymä, joka sisältää vähintään yhden leksikaalisesti spesifioidun elementin ko(n)tekstuaalisine valintoineen. Se esiintyy yleisemmin kuin yksittäistapauksissa, ja sen osat voivat esiintyä samassa virkkeessä joko peräkkäin tai jonkin välimatkan päästä toisistaan. Se sallii jossain määrin leksikaalista ja syntaktista joustavuutta ja toimii yhtenä semanttisena yksikkönä, mutta se ei ole välttämättä semanttisesti epäkompositionaalinen.

3.2.3. Keskeiset fraseologiset yksiköt ja käsitteet

Fraseologinen tutkimus on paljolti sanojen syntagmaattisten suhteiden (ks. luku 3.1.3.) tarkastelua. Nämä suhteet realisoituvat tekstissä joko tiukkoina valintarajoituksina tai väljempinä valintapreferensseinä, ja niiden taustalla toimii Sinclairin (1991) idiomiperiaate (ks. luku 3.1.4.), jonka mukaan sanojen käyttöä ohjaavat muun muassa lukuisat esiintymisrajoitukset ja -preferenssit (Jantunen 2009b: 358). Stubbs (2001a: 64–66, 87–88) on tunnistanut Sinclairin (1996; 1998) työn pohjalta neljä sellaista suhdetta, joita hän kutsuu myötäesiintymissuhteiksi (*co-occurrence relations*). Hän luettelee ne laajennettujen leksikaalisten yksiköiden mallissaan (*model of extended lexical units*):

1. **kollokaatio** (*collocation*) koskee yksittäisiä sanoja,
2. **kolligaatio** (*colligation*) koskee sanaluokkia,
3. **semanttinen preferenssi** (*semantic preference*) koskee sanajoukkoja, joilla on samanlaisia semanttisia piirteitä ja
4. **semanttinen prosodia** (*semantic prosody* tai *discourse prosody*) koskee vielä laajempaa semanttista ilmiötä ja sisältää paljon leksikaalista vaihtelua. (Stubbs mp.)

Suhteet etenevät nousevassa abstraktiusjärjestyksessä. Niistä ensimmäiset kaksi ovat selkeimmin havaittavissa, sillä ne ilmenevät tekstissä konkreettisina myötäesiintyminä. Jälkimmäiset kaksi ovat taas vaikeammin havaittavissa, sillä ne ovat pikemminkin abstrakteja semanttisia ominaisuuksia kuin konkreettisia ilmentymiä. Ero niiden välillä ei

myöskään ole täysin selkeä (Stubbs 2001a: 66). Kaikki käsitteet tulevat esiintymään analyysissäni, joten käyn ne läpi tarkemmin seuraavissa alaluvuissa.

3.2.3.1. Kollokaatio

Kollokaation nimitys on Partingtonin (1998: 15) mukaan peräisin Firthiltä (1957, 1968 [1957]), kuten myös alalla tunnetuksi tullut lause ”You shall know a word by the company it keeps!” (Firth 1968 [1957]: 179). Termi viittaa yksinkertaisimmillaan kahden sanan myötäesiintymiseen, mutta sen tarkemmassa määrittelyssä vallitsee jonkin verran vaihtelua tutkijoiden keskuudessa. Ainakin kolme suuntaa on eroteltavissa, nimittäin kollokaation tekstuaalinen, assosiatiivinen sekä tilastollinen määritelmä (ks. esim. Partington 1998; Jantunen 2001).

Tekstuaalinen määritelmä. Tämän määritelmän mukaan kollokaatio on kahden tai useamman sanan myötäesiintymä tekstissä toistensa läheisyydessä. Se koostuu hakusanasta eli **noodista** (*node*) ja sen **kollokaateista** (*collocates*), jotka esiintyvät noodin yhteydessä ennalta määrätyn välimatkan eli **kollokaativälin** (*span*) sisällä. Tämä kollokaativäli vaihtelee tutkijoittain, mutta se on yleensä korkeintaan neljä sanaa noodin molemmin puolin. (Sinclair 1991: 170.) Jantunen (2001: 173) selventää, että määritelmän mukaan hakusanan kollokaatti voi olla mikä tahansa sana, joka esiintyy hakusanan kanssa kollokaativälin sisällä, eikä kollokaation osien tarvitse sijaita tekstissä vierekkäin. Vaikka käytänkin tätä määritelmää omassa tutkimuksessani, en määrittele tässä tarkkaa kollokaativäliä, koska pystyn helposti tarkastelemaan kollokaatioita virkkeen laajuudessa kontekstissa ICLFIn raakateksteistä.

Assosiatiivinen määritelmä. Tässä määritelmässä kyse on merkitykseltään yhteen liittyvien sanojen taipumuksesta esiintyä toistensa yhteydessä (Leech 1981: 17; Karlsson 2008: 232). Leech viittaa tähän määritelmään puhuessaan kollokatiivisesta merkityksestä (*collocative meaning*) ja antaa esimerkiksi englanninkielisen lähisynonyymiparin *pretty* ’nätti’ ja *handsome* ’komea’. Vaikka kumpikin sana viittaa hyvännäköisyyteen, *pretty* luonnehtii useimmiten naispuolisia ja feminiiniseksi mielletäviä tarkoitteita, kun taas *handsome* luonnehtii useimmiten miespuolisia ja maskuliiniseksi mielletäviä

tarkoitteita. Leechin mukaan kollokatiivisessa merkityksessä on kyse pikemminkin yksittäisten sanojen idiosynkraattisista ominaisuuksista kuin yleistettävissä olevista säännöistä.

Karlssonin (2008: 232–233) mukaan kollokaation taustalla on tarkoitteiden luontainen yhteen kuuluminen, ja hän antaa esimerkiksi verbit AMMUA ja HIRNUA, joiden subjekti-tarkoitteet rajoittuvat lehmiin (AMMUA) tai hevosiin (HIRNUA). Pirkola (2016: 11) kuitenkin huomauttaa, että Karlssonin väite ei päde sellaisiin sanoihin, joilla ei ole kielenulkoista tarkoitetta, kuten Pirkolan tutkimat EHKÄ ja MAHDOLLISESTI. Jantunen (2009b: 358) toteaaakin, että kollokaatio voi usein motivoitua muusta kuin kielenulkoisesta maailmasta. Noudatan itse tätä laajempaa käsitystä, kun käsittelen kollokaatiota assosiatiivisesta näkökulmasta.

Tilastollinen määritelmä. Keskeistä tässä määritelmässä on toistuvuus: ollakseen tilastollisesti merkittävä kollokaation täytyy toistua tutkittavassa aineistossa riittävän usein (Jantunen 2001: 173). Tätä määritelmää edustaa esimerkiksi Hoey (1991: 6–8), jonka mukaan kollokaatio on leksikaalisen yksikön suhde muihin yksikköihin, jotka esiintyvät sen kanssa useammin kuin satunnaisesti yksikön (tekstuaalisessa) kontekstissa. Myös Biber ja Conrad (1999: 183) pitävät kollokaatiota pikemminkin tilastollisina assosiaatioina kuin suhteellisen kiinteinä ilmauksina, joissa yksittäisten sanojen merkitykset eivät säily. He kuitenkin huomauttavat, että sanan laajennettuun merkitykseen kuuluu myös sen taipumus esiintyä tiettyjen kollokaattien kanssa.

Tilastollista määritelmää suosii myös Stubbs (2001a: 29), joka määrittelee kollokaation frekventiksi myötäesiintymiseksi (*frequent co-occurrence*). Hän nostaa esille korpuslingvistiikassa vallitsevan oletuksen, että merkittäviä ovat nimenomaan frekventit tapaukset. Partington (1998: 16) mainitseekin, että juuri kollokaation tilastollinen määritelmä soveltuu hyvin korpuslingvistiseen tutkimukseen, jossa suuria aineistomääriä pystytään analysoimaan tietokoneohjelmien avulla. Vaikken käytäkään tilastollisia testejä omassa analyysissäni, pystyn saamaan hyvän kuvan sanojen yleisimmistä kollokaateista tutkimalla niiden konkordanssituloksia.

Nojaudun tässä tutkimuksessa kollokaation kaikkiin kolmeen määritelmään. Pirkolan (2016: 12) tapaan en pidä niitä ristiriitaisina tai toisiaan pois sulkevinä, ja ottamalla kaikki

ulottuvuudet huomioon toivon saavani mahdollisimman kokonaisvaltaisen kuvan edistyneessä oppijansuomessa esiintyvistä kollokaatioista.

3.2.3.2. Kolligaatio

Siinä missä kollokaatio on puhtaasti leksikaalinen myötäesiintymä, **kolligaatiossa** on kyse leksikaalis-kieliopillisesta myötäesiintymästä eli sanan ja kieliopillisen kategorian välisestä syntagmaattisesta suhteesta (Sinclair 1998: 15; Stubbs 2001b: 449; Jantunen 2009b: 358). **Kolligaatti** taas on se kielioppikategoria, jonka yhteydessä hakusanalla on taipumus esiintyä. Nämä kieliopilliset kategoriat voivat olla esimerkiksi sanaluokka tai lauseenjäsen. Myös suomenoppijoille vaikea rektio-ilmiö, jossa valitaan sanan täydennysten sijamuodot, kuuluu kolligaation piiriin. (Jantunen 2001: 174; 2004: 22; 2009b: 359.)

Kolligaatio on kollokaatiota abstraktimpi käsite. Siinä missä kollokoivat lekseemit ovat havaittavissa suoraan tekstistä, kolligaation havaitseminen vaatii puolestaan sitä, että kunkin tarkasteltavan kohteen sanaluokka on tiedossa. (Sinclair 1998: 16.) Jantunen (2009b: 359) havainnollistaa kolligaatiota antamalla kaksi esimerkkiä astemääritteen KOVIN esiintymist taipumuksista. Hänen mukaansa KOVIN tapaa esiintyä muun muassa kieltoverbin yhteydessä, ja lisäksi se esiintyy useammin kvanttorien (esim. HARVA, PALJON) seurassa kuin sen synonyymiset astemääritteet HYVIN ja OIKEIN. Tämä voi olla syy siihen, miksi aineistoni ilmaus *Aleksanteri II:n aika oli kovin tärkeä ja edullinen Suomelle* (PU0007c) saattaa kuulostaa natiivin mielestä oudolta.

Muun muassa Halliday (1966: 159) ja Jantunen (2001: 174) ovat puoltaneet leksikaalis-kieliopillisten suhteiden tarkastelua kollokaatioiden rinnalla, sillä sanat voivat esiintyä myös vakiintuneissa kieliopillisissa rakenteissa. Leksikaalis-kieliopillisten suhteiden tarkastelu onkin osoittautunut hyväksi apukeinoksi synonyymisten sanojen ja rakenteiden erottelussa (Biber, Conrad & Reppen 1998: 84). Koska aineistossani esiintyy paljon lähisynonyymien epätyypillistä käyttöä ja haluan selvittää, toimivatko ne erilaisissa kieliopillisissa tehtävissä kuin natiivisuomessa, otan kolligaation tarkastelun mukaan tutkimukseeni.

3.2.3.3. Semanttinen preferenssi

Siinä missä kollokaatio on leksikaalinen myötäesiintymä ja kolligaatio on leksikaalis-kieliopillinen myötäesiintymä, **semanttinen preferenssi** on puolestaan leksikaalis-semanttinen myötäesiintymä. Ilmiö viittaa siihen, että hakusanan yhteydessä säännönmukaisesti esiintyvät sanat jakavat jonkin yhteisen merkityspiirteen. Tämä merkityspiirre voi ilmetä eri positioissa hakusanan kontekstissa, eivätkä sitä ilmentävät sanat ole sidoksissa mihinkään sanaluokkiin. (Sinclair 1998: 16; Jantunen 2004: 24; 2009b: 359.)

On huomattava, että tämä leksikaalis-semanttinen suhde ei esiinny yksittäisten sanojen välillä, vaan kyseessä on hakusanan ja semanttisesti yhtenäisen sanajoukon välinen suhde. Nämä sanajoukot ovat abstrakteja ja avoimia sen lisäksi, että ne sisältävät joukolle tyyppisimpiä jäseniä. (Stubbs 2001a: 65, 88.) Jantunen (2009b: 359) esimerkkiä lainaten voisi sanoa, että kun astemääritteellä OIKEIN on semanttinen preferenssi 'arvoa' tai 'hyvyyttä' merkitseviin sanoihin kuten HYVÄ, HYVIN ja MAINIO, suhde on siis sanan OIKEIN ja sanajoukon (johon kuuluvat mm. HYVÄ, HYVIN, MAINIO) välillä. Kun OIKEIN-sanaa käytetään sanajoukkoon kuulumattoman sanan yhteydessä, kuten aineistoni esimerkissä *Hän on oikein onneton* (VE0100), ilmaus saattaa kuulostaa natiivin mielestä oudolta.

Semanttinen preferenssi on Bednarekin (2008: 122) mukaan kollokationaalinen ilmiö, joka pystytään suhteellisen objektiivisesti havaitsemaan korpusaineistosta. Hän kuitenkin huomauttaa, että merkityspiirteiden määrittely sekä kollokaattien semanttinen tulkinta ja luokittelu riippuu paljolti tutkijasta, ja erityisesti kollokaattien objektiivinen arviointi positiiviseksi tai negatiiviseksi voi osoittautua vaikeaksi. Lisäksi sanoilla voi olla useammanlaisia semanttisia preferenssejä eri merkitysryhmiin (Jantunen 2004: 25) tai erilaisia semanttisia preferenssejä kirjaimellisissa ja kuvaannollisissa käyttöyhteyksissä (Bednarek 2008: 125). Näitä hienovaraisuuksia ei Bednarekin mukaan välttämättä tule ilmi korpusaineiston automaattisessa analyysissä.

3.2.3.4. Semanttinen prosodia

Semanttinen prosodia on semanttisen preferenssin tavoin leksikaalis-semanttinen myötäesiintymä, mutta se on semanttista preferenssiä abstraktimpi (Jantunen 2004: 25; 2009b:

360). Se viittaa tietyllä tekstialueella vallitsevaan konnotatiiviseen merkityspiirteeseen, joka ylittää yksittäisten sanojen rajat (Partington 1998: 67–68; 2004: 131–132). Louw (1993) määrittelee semanttisen prosodian tarkasteltavaa sanaa ympäröiväksi merkitysauraksi (*aura of meaning*).

Semanttista prosodiaa on usein kuvattu käsiteparilla *negatiivinen–positiivinen*, jossa hakusanan negatiivista tai positiivista semanttista prosodiaa heijastavat sen kollokaattien kielteiset tai positiiviset merkityssisällöt (Jantunen 2004: 26). Esimerkiksi semanttinen prosodia on useimmiten positiivinen sanalla *elämys* ja negatiivinen sanalla *joutua*, minkä takia aineistoni ilmaukset *kauheita [elämyksiä]* (PU0032f) ja *[elokuva joutui valkokankaalle]* (VI0365a) saattavat kuulostaa natiivin mielestä oudolta. Toisinaan semanttinen prosodia voi myös olla neutraali (Jantunen 2009b: 359).

Semanttinen prosodia määritellään myös ilmauksen pragmaattiseksi merkitykseksi, joka heijastaa kielenkäyttäjän asenteita ja josta ilmenevät ne syyt siihen, miksi juuri kyseinen ilmaus on valittu (Sinclair 1996: 87–88, 1998: 20). Tätä näkemystä puoltaa Stubbs (2001a: 65–66; 2001b: 449), joka onkin esittänyt ilmiölle vaihtoehdoisen termin **diskurs-siprosodia** (*discourse prosody*) korostaakseen sen vuorovaikutteisuuutta ja tilannesidon-naisuutta. Partington (2004: 132) puolestaan ehdottaa, että ilmiö selittynee Hoeyn (mm. 2005) priming-käsitteellä (ks. luku 3.1.5.), sillä semanttiset prosodiat eivät kumpua itse sanojen ”synnynnäisistä” merkityksistä, vaan ne pohjautuvat kielenkäyttäjien primingeihin, jotka säätelevät eri sanojen käyttöympäristöjä.

Semanttinen prosodia on hyvin kontekstisidonnainen ilmiö, jonka tulkinta vaatii laajempaa tekstiyhteyttä kuin edellä mainitut syntagmaattiset suhteet (Jantunen 2004: 27). Sen tutkiminen on usein hankalaa, sillä ilmiö ei ole suoraan nähtävissä tekstistä, vaan tutkija joutuu käyttämään kieli-intuitiotaan sen havaitsemiseksi (Bednarek 2008: 132). Analysoinnin vaikeutta havainnollistaneen Partingtonin (2004: 154–155) huomautus siitä, että hakusanan kollokaattien semanttiset piirteet eivät välttämättä heijasta hakusanan semanttista prosodiaa. Esimerkiksi LIEVITTÄÄ, jolla on lähtökohtaisesti positiivinen merkitys, esiintyy usein negatiivisiin tarkoitteisiin viittaavien sanojen kuten SÄRKY, KIPU ja OIRE yhteydessä. Negatiivisista kollokaateistaan huolimatta LIEVITTÄÄ-lekseemillä voidaan sanoa olevan positiivinen semanttinen prosodia, sillä se viittaa huonon tilanteen parantamiseen.

Semanttiset prosodiat ovat monesti hienovaraisia ja vaihtelevia (Sinclair 1998: 20), ja niiden laatu ja vahvuus saattavat hyvinkin vaihdella genreittäin ja käyttöalueittain (Partington 2004: 153). Louw (1993: 157, 172–173) mainitsee, että semanttiset prosodiat ovat olleet pitkälti introspektion tavoittamattomissa, eikä niitä pystytä helposti löytämään edes korpusaineistoista. Hän (mts. 164) myös toteaa, että jos halutaan saada tarkka ja luotettava tieto semanttisesta prosodiasta, on syytä koota entistä isompia korpuksia. Tällainen tieto on Partingtonin (1998: 72) mukaan erityisen tärkeä kielenoppijoille, jotka eivät välttämättä ole yhtä perillä kielen vivahteista kuin äidinkielliset kielenkäyttäjät.

3.3. Oppijankieli ja edistynyt kielitaito

3.3.1. Oppijankielen käsite ja luonne

Kieltä voi oppia kahdella tavalla: sitä voi *omaksua toisena kielenä* (*second language acquisition*) kieliyhteisön parissa tai *oppia vieraana kielenä* (*foreign language learning*) kieliyhteisön ulkopuolella. On kuitenkin mahdotonta tehdä täydellistä jakoa toisen kielen omaksumisen ja vieraan kielen oppimisen välillä, vaan kyseessä on pikemminkin jatkumo, jonka ääripäissä kumpikin oppimiskonteksti sijaitsee. (Nissilä 2011: 37–38.) Latomaan ja Tuomelan (1993: 243) mukaan puhtaimmillaan suomi toisena kielenä olisi kielimuoto, jota omaksutaan Suomessa yksinomaan ”kadulla”, kun taas suomi vieraana kielenä olisi kielimuoto, jota opitaan vain opetuksen kautta ulkomailla.

Kielenoppija on siis henkilö, joka omaksuu tai oppii toista tai vierasta kieltä. **Oppijankieli** viittaa taas kielimuotoon, jonka kielenoppija tuottaa. Oppijankielestä on käytetty myös Selinkerin (mm. 1969, 1972) lanseeraamaa **välikieli**-termiä (*interlanguage*), mutta nimitys on toisinaan koettu ongelmalliseksi, sillä siihen saattaa sisältyä oletus oppijankielen puutteellisuudesta ja vajavaisuudesta (Martin & Alanen 2011: 34). Suosin tästä syystä neutraalimpaa *oppijankieli*-termiä tässä tutkielmassa.

Kun puhutaan kielenoppijan äidinkielestä, käytetään termiä **lähtökieli**, **lähdekieli** tai **ensikieli**, ja kun puhutaan opeteltavasta kielestä, käytetään puolestaan **kohdekieli**-termiä. Äidinkielliseen puhujaan viitataan nimityksellä **natiivi** (*native speaker*), kun taas toisen tai vieraan kielen puhujasta käytetään nimitystä **ei-natiivi** (*non-native speaker*). (Nissilä

2011: 38.) Suomen kielen kontekstissa **natiivisuomi** on se suomen kielen muoto, jota äidinkielen kielenkäyttäjä tuottaa. **Suomenoppija** on henkilö, joka on joko omaksunut suomea toisena kielenä tai oppinut sitä vieraana kielenä. Tämän tutkimuksen suomenoppijat ovat oppineet suomea pääosin vieraana kielenä ulkomaisissa yliopistoissa, ja juuri heidänlaisiinsa viitataan, kun käytän tutkielmassani termiä *suomenoppija*. Vastaavasti **oppijansuomi** voi tarkoittaa joko suomea toisena kielenä omaksuneen tai vieraana kielenä oppineen suomenoppijan tuottamaa suomea. Tässä tutkielmassa *oppijansuomi* viittaa kuitenkin siihen kielimuotoon, jota nimenomaan suomi vieraana kielenä -oppijat tuottavat.

Oppijankieltä pidetään itsenäisenä lähtö- ja kohdekielen välisenä järjestelmänä, johon vaikuttavat muun muassa kielenoppijan lähtökieli, kohdekieli, meneillä oleva oppijankielen vaihe sekä muut kielenoppijan osaamat kielet. Se on systemaattinen mutta myös dynaaminen ja luova vaihe kielenoppimisessa, jossa vanhoja sääntöjä muokataan ja uusia omaksutaan koko ajan. Tämä jatkuvasti muuttuva luonne tekee kielenoppijan oppimiskulun ennustamisen vaikeaksi, mutta siitä huolimatta yhteisiä piirteitä eri oppijoiden välillä on löydettävissä. (Nissilä 2011: 39–40.)

Jantunen (2008) on esittänyt hypoteeseja ominaispiirteistä, joita esiintyy kaikissa oppijankielissä riippumatta kielenoppijan lähtökielestä, kohdekielestä tai muista osaamista kielistä. Näihin **oppijankielen universaaleihin** kuuluvat Jantunen mukaan kielenainesten epätyypilliset frekvenssit (*atypical frequencies of linguistic items*), kieltenvälinen vaikutus (*crosslinguistic influence*), kielellinen yksinkertaisuus (*simplicity*) sekä epäkonventionaalisuus (*unconventionality*), joka puolestaan jakautuu yleis- ja puhekielen ”normistojen” sekoittumiseen (*standard language vs. spoken language*) ja kontekstuaaliseen epäkonventionaalisuuteen (*contextual unconventionality*).

Yllä mainituista oppijankielen universaaleista juuri kontekstuaalisen epäkonventionaalisuuden käsite on keskeinen tässä tutkimuksessa. Kontekstuaalisella epäkonventionaalisuudella Jantunen (2008: 15) tarkoittaa ”tekstuaalista, kontekstuaalisen merkitysyksikön rakenteellista epätyypillisyyttä” eli fraseologista epätyypillisyyttä, joka koskee muun muassa kollokaatiota, kolligaatiota, semanttista preferenssiä ja semanttista prosodiaa. Äidinkielen kielenkäyttäjän kompetenssiin kuuluu Jantunen mukaan tieto kielen ilmausten tyypillisestä tai prototyyppisestä käytöstä, joskin nämä säännöt ja preferenssit eivät ole aina tarkasti selitettävissä tai eriteltävissä edes natiivillakaan. Äidinkielen

kielenkäyttäjän tuotokset noudattavat siis pääosin Sinclairin (1991) idiomiperiaatetta, kun taas kielenoppija tuottaa kieltä vapaan valinnan periaatteen mukaan silloin, kun häneltä puuttuu äidinkielen kielenkäyttäjän tieto ilmausten tyypillisestä käytöstä. Käytännössä kielenoppija joutuu siis arvaamaan, mikä kielellinen yksikkö sopii mihinkin ko(n)tekstiin. Tämä arvaaminen johtaa epäkonventionaalsiin valintoihin, jotka puolestaan ilmenevät fraseologisina epätyypillisyyksinä tekstissä. (Jantunen 2008: 16.)

3.3.2. Virheet ja epätyypillisyydet

Virheitä on tavallisesti pidetty asiana, joka erottaa äidinkielisten kielenkäyttäjien tuotokset kielenoppijoiden tuotoksista (Martin 1995: 264). Virheiden tarkastelu onkin muodostanut merkittävän osan oppijankielen tutkimusta kautta sen erilaisten vaiheiden (ks. esim. Larsen-Freeman & Long 1991: 52–74). Erityisesti virheanalyysin (ks. esim. Corder 1973, 1974, 1981; James 1998; Ellis & Barkhuizen 2005: 51–71; myös Sajavaara 1999: 104, 116–119) kukoistuskautena panostettiin virheiden tunnistamiseen, kuvaamiseen, selittämiseen sekä arviointiin. Virheet tarjoavat Corderin (1967: 167) mukaan kielenoppijan käyttämään kielijärjestelmään liittyvää tietoa, jota opettajat, tutkijat sekä itse kielenoppijatkin voivat hyödyntää.

Niin kielenoppijat kuin äidinkielliset kielenkäyttäjätkin tekevät virheitä (Martin 1995: 264). Kuitenkin Corderin (1967; 1973: 256, 257) mukaan suurin osa kielenoppijoiden tekemistä virheistä eroaa perusluonteisesti niistä virheistä, joita äidinkielliset kielenkäyttäjät tekevät. Hän (1967: 166–167) jakaa virheet epäsystemaattisiin **performanssivirheisiin** (*errors of performance*), jotka syntyvät esimerkiksi muistikatkon, väsymyksen tai voimakkaiden tunteiden vaikutuksesta, sekä systemaattisiin **kompetenssivirheisiin** (*errors of competence*), jotka viittaavat kielenkäyttäjän puutteelliseen tietoon käyttämästään kielen järjestelmästä. Edellisestä virhetyypistä Corder käyttää englanninkielistä termiä *mistake* ja jälkimmäisestä virhetyypistä termiä *error*. Äidinkielisten kielenkäyttäjien tekemät virheet ovat Corderin (1967, 1973) mukaan lähinnä performanssivirheitä, kun taas kompetenssivirheitä tekevät enimmäkseen kielenoppijat. Lisäksi hän (1973: 260–261) mainitsee **lipsahdukset** (*lapses*), joita tekevät sekä äidinkielliset kielenkäyttäjät että kielenoppijat ja jotka eivät ole kovin olennaisia kielenoppimisen kannalta.

Virheitä on myös pyritty selittämään erityisesti virheanalyysissa. Richards (1974: 173) kutsuu **interlingvaalisiksi virheiksi** sellaisia virheitä, jotka aiheutuvat kielenoppijan äidinkielestä ja **intralingvaalisiksi virheiksi** sellaisia virheitä, jotka ovat yleisiä kielenoppijoiden keskuudessa heidän äidinkielestään huolimatta. Yhtenä virheiden aiheuttajana pidetään **kielten välistä vaikutusta** (*crosslinguistic influence*) eli **siirtovaikutusta** (*transfer*), tarkemmin **negatiivista siirtovaikutusta** (*negative transfer*) eli **interferenssiä** (*intereference*) (ks. esim. Odlin 1989; R. Ellis 2008: 349–403). Sajavaara (1999: 116) lisää virhetyyppien joukkoon myös **opetuksen aiheuttamat virheet** ja toteaa, ettei virheiden selittäminen ole aina helppoa.

Virheanalyysissa onkin havaittu useita ongelmia oppijankielen tutkimuksessa. Menetelmää on kritisoitu muun muassa siitä, että sen avulla ei pystytä saamaan kokonaiskuvaa oppijankielestä. Lisäksi virheet eivät ole aina helposti luokiteltavissa ja niiden yksiselitteisen lähteen löytäminen on monesti vaikeaa tai jopa mahdotonta. (Larsen-Freeman & Long 1991: 61–62; Latomaa 1995: 80; ks. myös Schachter & Celce-Murcia 1977.) Latomaa (mp.) mainitsee, että virheiden on todettu kuuluvan kielenoppimisprosessiin eikä niitä tarvitse tai edes voi ehkäistä. Nissilä (2011: 39–40) puolestaan esittää, että oppijankielessä esiintyviä poikkeamia kohdekielestä ei pidetä virheinä vaan oppijankielen ominaisuuksina.

Vaikka yhdyinkin näkemykseen, että kohdekielen normeista poikkeaviin tuotoksiin ei tulisi suhtautua kielteisesti, vedän tässä tutkimuksessa rajan virheiden ja epätyypillisyyksien välillä. Kutsun **virheiksi** sellaisia tapauksia, jotka rikkovat selvästi jotakin kohdekielen kielioppisääntöä. Tällainen on esimerkiksi transitiivi- ja intransitiiviverbin sekoittaminen lauseessa **Jacek **aloitti** elää rikkaana hallitsijana*, – – (PU0005e). Kutsun **epätyypillisyyksiksi** puolestaan tapauksia, jotka ovat Jantusen (2008: 15–18) sanoin kontekstuaalisesti epäkonventionaalisia, mutta kuitenkin kieliopillisesti virheettömiä ja semanttisesti yksiselitteisesti ymmärrettäviä. Tällainen on esimerkiksi epätyypillisen verbin käyttäminen lauseessa *?Kun tilanne **pilaantui***, – – (PU0005k). Keskityn analyysissani lähinnä epätyypillisyyksiin, mutta nostan esille myös virheitä, joita pidän huomionarvoisina edistyneeseen kielitaitotasoon nähden.

3.3.3. Kielitaito ja Eurooppalainen viitekehys

Oppijankielen kehitystä voidaan tarkastella muun muassa tarkkuuden, sujuvuuden ja kompleksisuuden näkökulmasta, joskin käsitteiden määrittely ja mittaaminen voivat olla vaihtelevaa ja ongelmallista. **Tarkkuus** viittaa virheiden puuttumiseen kielenoppijan tuotoksissa. Koska kohdekielen normi vaihtelee tehtävien mukaan, se on syytä määritellä mahdollisimman tarkasti tutkimuksessa. **Sujuvuuden** tieteellinen määrittely on hankalaa, vaikka maallikolle käsite lienee selvä. Sen arvioinnissa otetaan huomioon muun muassa kielenoppijan tiettyssä ajassa tuottama ilmaus-, sana- tai morfeemimäärä sekä taukojen ja korjausten määrä. **Kompleksisuus** tarkoittaa puolestaan sitä, kuinka mutkikkaita kielenoppijan tuotokset ovat kielellisesti ja kognitiivisesti. Yleensä sitä mitataan määrällisillä mittareilla. (Martin & Alanen 2011: 39.)

Kielenoppijoiden kielitaidon kuvaamiseksi käytetään erilaisia arviointiasteikkoja, joista yksi on **Eurooppalainen viitekehys** (*Common European Framework of Reference for Languages*) eli **EVK**. Se kuvaa taitotasoinnista sitä, mitä kielenoppijoiden tulisi oppia sekä mitä tietoja ja taitoja heidän tulisi kehittää, jotta he pystyisivät käyttämään kohdekieltään tehokkaasti viestintään. EVK:n arviointiasteikko jakautuu kuuteen taitotasoportaaseen. Perustasoon kuuluvat A1- eli alkeistaso (*breakthrough*) ja A2- eli selviytyjän taso (*way-stage*), kun taas keskitaso koostuu B1- eli kynnystaitotasosta (*threshold*) ja B2- eli osaajan tasosta (*vantage*). Edistyneen tason muodostavat puolestaan C1- eli taitajan taso (*effective operational proficiency*) sekä C2- eli mestarin taso (*mastery*). (EVK 2003: 19, 46–47.)

Keskityn tässä tutkimuksessa fraseologisiin epätyypillisyyksiin, jotka esiintyvät edistyneen eli C1- ja C2-tason teksteissä. EVK:n eri tarkoituksiin kehitetyistä taitotasokuvauksista tutkimukseni kannalta olennaisimmat liittyvät kielenoppijoiden sanastolliseen ja kieliopilliseen kompetenssiin. **Sanastollinen kompetenssi** tarkoittaa ”kielen sanaston ja kieliopin elementeistä koostuvan sanavaraston tuntemusta ja kykyä käyttää sitä”. Edistyneen kielenoppijan sanastolliseen kompetenssiin kuuluu muun muassa laajan sanaston sekä kohdekielelle tyypillisten idiomaattisten ilmausten hyvä hallinta. Lisäksi hän kykenee käyttämään sanastoa asianmukaisesti eikä tee erityisen vakavia sanastovirheitä (jos niitä tekee ylipäättään). **Kieliopillinen kompetenssi** tarkoittaa puolestaan ”kielen kieliopillisten keinojen tuntemusta ja kykyä käyttää näitä keinoja”. Edistyneen kielenoppijan kieliopilliseen kompetenssiin kuuluu lähes kaikkien – myös monimutkaisten – kielioppi-

rakenteiden johdonmukainen ja virheetön hallinta. Jos kielioppivirheitä esiintyy, ne ovat harvinaisia ja lähes huomaamattomia. (EVK 2003: 157–163.)

Mikään kielitaidon arviointimenetelmä ei ole täysin ongelmaton, ja EVK:stä onkin esitetty kritiikkiä (ks. esim. Little 2007, Alderson 2007). Pysin analyysissäni selvittämään, kuinka hyvin todellisissa kielenkäyttötilanteissa tuotetut ja taitotasoarvioidut edistyneen oppijansuomen tekstit vastaavat EVK:n kuvauksia ainakin yllä mainituilla kielitaidon osa-alueilla.

3.3.4. Edistyneen kielenoppijan fraseologiset haasteet

Kuten edellisestä alaluvusta käy ilmi, edistyneen oppijankielen ominaisuuksiin pitäisi kuulua erittäin hyvä – ellei jopa virheetön – sanastollinen ja kieliopillinen hallinta. Kuitenkin käytännössä edistyneilläkin kielenoppijoilla voi olla suuria vaikeuksia natiivinkaltaisen idiomaattisen kielen tuottamisessa (Nesselhauf 2003: 237; N. Ellis 2008: 7). Esimerkiksi Ivaska (2015: 3–4) mainitsee, että edistyneet kielenoppijat pystyvät toimimaan kohdekielellään akateemisen ympäristön kaltaisissa tilanteissa, jotka vaativat monipuolista ja abstraktia kielitaitoa. Tutkimuksessa on kuitenkin hänen mukaansa havaittu, että vaikka kielenoppijat osaavat usein kohdekieltään laadullisesti hyvin, heidän kielelliset valintansa ovat käyttökontekstissaan epätyypillisiä.

Fraseologian puutteellinen hallinta näyttää siis olevan kompastuskivi edistyneille kielenoppijoille, jotka tavoittelevat johdonmukaisesti natiivinkaltaisia tuotoksia kohdekielellään. Laajoista aineistoista onkin ilmennyt, että kollokaatiot ovat yksi vaikeimmin opittavissa olevista asioista, joita kielenoppijat kohtaavat sanastoa oppiessaan. Syy tähän on kollokaatioiden arbitraarisuus, joka pakottaa kielenoppijat opettelemaan kunkin yksittäisen rakenteen ulkoa. Kun rakenteita on miltei rajattomasti, ei liene ihme, että epätyypillisiä kollokaatioita esiintyy yhä edistyneidenkin kielenoppijoiden tuotoksissa. (Jantunen 2009b: 361.)

Fraseologisten yksiköiden ja ilmiöiden arbitraarisuus tarkoittaa myös sitä, ettei kielenoppijoilla ole käytössään kielioppisääntöjen kaltaisia valmiita ohjeita siihen, miten sanoja tulisi sijoittaa erilaisiin opittuihin rakenteisiin. Sen sijaan he joutuvat ikään kuin

kokeilemaan joka kerta, sopiiko heidän tuntemansa tai sanakirjasta löytämänsä sana oikeaan käyttöyhteyteen. Apua löytyy harvoin sanakirjasta, sillä useimmat sanakirjat esittävät sanat kontekstistaan irrallisina yksikköinä. Ongelma koskee erityisesti heitä, jotka oppivat kohdekieltään vieraana kielenä, sillä tällaisilta kielenoppijoilta puuttuu ympäröivän kieliyhteisön tarjoama malli. (Jantunen 2009b: 356.) Kun puutteelliset tai suorastaan olemattomat oppi- ja opetusmateriaalit tekevät fraseologian oppimisen ja opettamisen hankalaksi, kielenoppijoiden fraseologisen tuntemuksen kehitys saattaa jäädä kokonaan sattuman varaan (Lindroos-Čermáková 1993: 91–92).

3.4. Aiempi tutkimus

Suomen kielen fraseologinen korpustutkimus on toistaiseksi ollut melko vähäistä, olipa kyseessä natiivi- tai oppijansuomi. Yhtenä alan edelläkävijöistä voi pitää Jarmo Jantusta, joka on tutkinut muun muassa adjektiivien TÄRKEÄ ja KESKEINEN (2001), astemääritteiden HYVIN, KOVIN ja OIKEIN (2004) sekä adverbien AIVAN ja synonyymisen verbiparin HIEVAHTAA–LIIKAHTAA (2009b) fraseologiaa natiivisuomessa. Natiivisuomen fraseologinen korpustutkimus näyttääkin painottuvan synonyymian tarkasteluun: Jantusen edellä mainittujen tutkimusten jälkeen on tutkittu esimerkiksi lähisynonyymeja VUOKSI ja TAKIA (Suoraniemi 2011), adverbien TÄYSIN ja KOKONAAN kollokaatteja ja semanttista prosodiaa (Kangas 2018) sekä synonyymien EHKÄ ja MAHDOLLISESTI kollokaatteja ja semanttisia preferenssejä (Pirkola 2016).

Oppijankieltä on tähän mennessä tutkittu paljon indoeurooppalaisten kielten – varsinkin oppijanenglannin – osalta. Sen sijaan ei-indoeurooppalaisten oppijankielten, kuten oppijansuomen tutkimus on ollut suhteellisen vähäistä. Tutkimus on painottunut virheanalyysiin, josta on ongelmallisuudestaan huolimatta tullut oppijankielen tutkimuksen peruspiilari. Virhelähtöisen tutkimusasetelman suosio näkyy myös oppijankielen korpustutkimuksessa, joka on sekin lisääntymässä entisestään. (Jantunen 2008: 1–3.)

Painottuminen virheiden tarkasteluun ei ole kuitenkaan yhtä läsnä oppijansuomen tutkimuksessa kuin muualla (Jantunen & Brunni 2012: 72). Sen sijaan on usein keskitytty vertailemaan natiivi- ja oppijansuomen välisiä eroja sekä selittämään kielimuotojen välillä esiintyvää vaihtelua muun muassa kielitaitotasoin (Varrio 2014: 2). Myös

oppijansuomen luonne ja sen oppimisen kehitys ovat olleet suosittuja tutkimusaiheita (Akgül 2013: 14). Oppijansuomen korpustutkimus alkoi 2000-luvun alussa, ja sen tien-
viittana on toiminut parikymmentä vuotta vanhempi oppijanenglannin korpustutkimuk-
sen perinne (Jantunen & Brunni mp.). Sen yleistyessä myös oppijansuomen fraseologisia
korpustutkimuksia on alkanut viime vuosien aikana ilmestyä, joskin vielä melko niukasti.

Oppijansuomen fraseologisessa korpustutkimuksessa on sekä vertailtu oppijansuomea
natiivisuomeen että keskitytty pelkästään oppijansuomen kuvaamiseen. Natiivi- ja oppi-
jansuomen vertailevat fraseologiset korpustutkimukset ovat olleet muun muassa PALJON-
adverbin kollokointi (Kallioranta 2009), frekventtien adjektiivien kollokaatiot (Akgül
2013), SANOJA-verbin ydin ja täydennysympäristö (Varrio 2014) sekä SAADA-verbin fra-
seologia (Tarvainen 2018).

Pelkästään oppijansuomea tarkastelevia fraseologisia korpustutkimuksia ovat muun mu-
assa oppijansuomen kolligaatit ketjuuntuviissa verbirakenteissa (Seppälä 2012), edisty-
neen oppijansuomen konstruktiopiirteet (Ivaska 2015), fraseologiset virheet ICLFIssä
(Kuuluvainen 2015) sekä sanojen HIRVEÄ ja KAUHEA käyttö ja merkitykset oppijan-
suomessa (Ikävalko 2019). Kuuluvaisen tutkimuksella onkin suuri vaikutus omaan tutki-
mukseeni, sillä hänkin tarkastelee ja luokittelee ICLFI:n oppijansuomen fraseologisia epä-
tyypillisyyksiä (joita hän kutsuu *virheiksi*). Lisäksi hänen luokittelujärjestelmänsä on laa-
tuaan ensimmäinen ja toimii mallina oman luokittelujärjestelmäni muodostamisessa.
Tästä huolimatta tutkimuksemme eroavat ainakin kolmella olennaisella tavalla.

Ensinnäkin Kuuluvaisen aineiston tekstit sijoittuvat pääosin EVK:n taitotasolle A2–B1,
kun taas aineistoni rajoittuu vain C1- ja C2-tasoihin. Toiseksi Kuuluvaisen aineiston kir-
joittajat edustavat neljää äidinkieltä: hollantia, ruotsia, tšekkiä sekä viroa. Aineistoni kir-
joittajien äidinkielet ovat puolestaan hollanti, kiina, puola, ruotsi, saksa, slovakki, tšekki,
unkari, venäjä sekä viro, joten tarkastelemieni epätyypillisyyksien kirjo saattaa olla laa-
jempi. Kolmanneksi Kuuluvaisen aineisto koostuu niistä ICLFI:n virheannotoiduista teks-
teistä, jotka ovat valmistuneet maaliskuun 2015 alkupuoleen mennessä. Oma aineistoni
taas koostuu kaikista ICLFI:n C1- ja C2-tasoisista raakateksteistä, enkä huomioi virhean-
notointia ollenkaan. Näistä syistä rohkenen väittää, että tutkimukseni ei ole kopio Kuulu-
vaisen tutkimuksesta vaan pikemminkin täydennys sille.

4. ANALYYSI

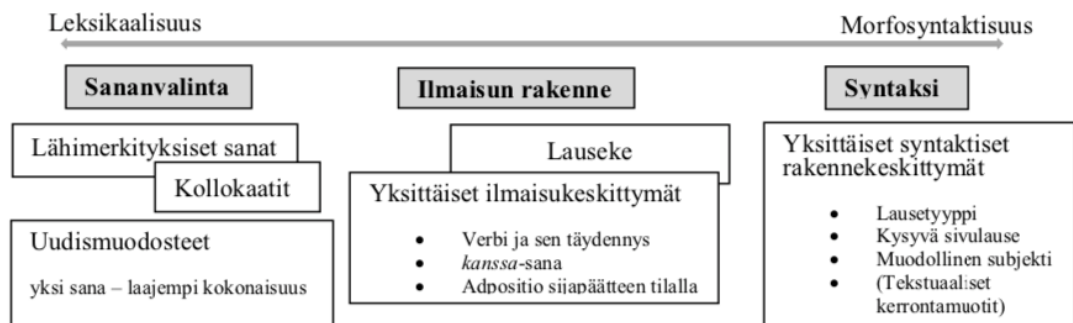
4.1. Epätyypillisyyksien valintaperusteet, luokittelu ja käsittely

Olen käyttänyt fraseologisten epätyypillisyyksien valinnassa omaa edistyneen suomenopijan kieli-intuitiotani. Alustavan valinnan perusteeksi on riittänyt tunne, että ilmaus ”kuulostaa jollain lailla oudolta”. Sen jälkeen olen käynyt valitsemani tapaukset läpi yhden äidinkielen kielenkäyttäjän kanssa, joka on myös suomen kielen ammattilainen. Selvitän lopulta epäilysteni paikkansapitävyyttä vertailuaineiston konkordanssituloksista, joiden avulla voin joko vahvistaa tai kumota hypoteesini.

Suoritan konkordanssihaut Kielipankin verkkopohjaisessa Korp-työkalussa. Hakuaineistona toimivat Ylen suomenkielisen uutisarkiston tutkijoille tarkoitettut osakorpukset vuosilta 2011–2018. **Hakusanalla** tai **-yhdistelmällä** tarkoitan sitä sanaa tai sanan joukkoa, jonka syötän yhdellä kerralla hakukenttään. Plusmerkillä erotettuja sanoja (esim. SANGEN + MUU) etsin samassa sanajärjestyksessä kuin ne kirjoitetaan, kun taas et-merkillä erotettuja sanoja (esim. PALATA & SEIKKA) etsin vapaassa sanajärjestyksessä aineistosta. Pyörästän hakuosumien luvut lähimpään kymmeneen, sillä tarkoitukseni ei ole tehdä tarkkoja tilastollisia analyyseja. Erityisesti suuremmat luvut (yli sata) ovat vain suuntaa antavia, sillä hakutuloksissa saattaa myös esiintyä pieni määrä hakukriteerejä vastaamattomia osumia. Jotta hakuni olisivat helposti replikoitavissa ja niiden tulokset helposti tarkistettavissa, olen luetellut kaikki käyttämäni CQP (*Corpus Query Protocol*) -kyselylausekkeet tämän tutkielman liitteessä.

Epätyypillisyyksien joukossa on niin kohdekielestä selkeästi poikkeavia kuin hienovaraisempiakin tapauksia. Edellisten tapausten joukkoon kuuluvat ne, joiden käyttöä määräävät suhteellisen tiukat, tilastollisesti merkittävät esiintymisrajoitukset. Tällaisia tapauksia ovat esimerkiksi kollokaatit ja idiomaattiset ilmaukset. Jatkumon toiseen päähän taas sijoittuu niitä sanoja, joiden käyttöä ohjaavat pikemminkin käyttöpreferenssit kuin -säännöt. Tällaisia tapauksia ovat esimerkiksi jotkin sananvalinnat, jotka saattavat vaihdella alueittain natiivisuomessakin. Toivon saavani tästä tutkimuksesta jonkinlaisen kuvan siitä, millaiset tapaukset ovat kohdekielestä selvästi poikkeavia ja millaiset tapaukset ovat taas tyyppillisyyden ja epätyypillisyyden rajalla.

Valmiina olevien vakiintuneiden luokittelutapojen puutteessa kehitin kandidaatintutkimukseni oman alustavan luokitteluni, jonka viisi eri luokkaa olivat epätyypilliset kollokaatit, kontekstiin sopimattomat sananvalinnat, epätyypillinen semanttinen prosodia, epätyypilliset lauserakenteet sekä epätyypilliset idiomaattiset ilmaukset (Cheng 2014). Hiljattain on kuitenkin ilmestynyt uusia oppijansuomen fraseologisia tutkimuksia, joista merkittävin tämän tutkimuksen kannalta on Kuuluvaisen (2015) ICLFIn fraseologisten virheiden luokittelu. Kuuluvainen kehitti omaa tutkimustaan varten leksikaalisuus–morfosyntaktisuus-jatkumon (ks. kuvio 1), johon hän sijoitti luokittelimiaan ICLFIn virheannotoituja tapauksia pää- ja alaryhmineen seuraavasti:



KUVIO 1. ICLFIn virheannotoidussa osakorpuksessa esiintyvien virheiden jatkumo (Kuuluvainen 2015: 31).

Kuuluvaisen luokittelu toimii runkona oman luokittelujärjestelmäni uudistamisessa. Otan käyttöön hänen perusideansa siitä, että tapaukset sijoittuvat leksikaalisuus–morfosyntaktisuus-jatkumolle ja jakautuvat sen perusteella eritasoisin pääryhmiin. Koska aineistomme koostuvat eritasoisten suomenoppijoiden tuotoksista, alaryhmiemme jaottelussa on odotettavissa suuriakin eroja. Säilytän vain osan Kuuluvaisen alaryhmistä omassa luokittelussani ja lisään siihen uusia alaryhmiä, jotka ovat muodostuneet analyysin varrella.

Olen päätenyt jakamaan löytämäni fraseologiset epätyypillisyydet neljään pääryhmään, jotka ovat sananvalinnan, morfologian, lauseketason sekä lausetason epätyypillisyydet. Pääryhmät jakautuvat useisiin alaryhmiin, jotka puolestaan saattavat vielä jakautua eri keskittymiin. Jako on vain suuntaa antava, sillä raja ryhmien välillä ei ole aina selkeä ja jossain kohdissa esiintyy päällekkäisyyksiä.

4.2. Sananvalinnan epätyypillisyydet

Sananvalinnan epätyypillisyyksien pääryhmään kuuluvat tapaukset, joissa yksi (epätyypillisempi) sana voidaan korvata toisella (tyypillisemmällä) sanalla muuttamatta lauseen muita osia ollenkaan. Vaihdettavat sanat ovat siis useimmissa tapauksissa paradigmaattisessa suhteessa (ks. luku 3.1.3.) toistensa kanssa, sillä ne viittaavat samantyyppiseen tarkoitteeseen. Kunkin sanan omat käyttörajoitukset ja -preferenssit ovat kuitenkin johtaneet siihen, että yksi sana sopii lauseen ko(n)tekstiin paremmin kuin toinen.

Tämä pääryhmää vastaa Kuuluvaisen (2015) luokittelun *sanavalintavirheiden* pääryhmää. Siitä hän erottaa kolme alaryhmää, jotka ovat *lähimerkityksisissä sanoissa esiintyvät virheet*, *kollokaattivirheet* sekä *uudismuodostesanat* eli suomenoppijoiden itse keksimät muodot, jotka ovat semanttisesti ymmärrettäviä mutta jollain tavalla natiivisuomesta poikkeavia. Omassa luokittelussani sananvalinnan epätyypillisyyksien pääryhmä jakautuu kahdeksaan alaryhmään.

Ensimmäiset neljä alaryhmää sisältävät pääosin epätyypillisyyksiä, jotka esiintyvät **yksittäisissä lähimerkityksisissä sanoissa, intensiteettipartikkeleissa, muissa kollokaatioissa ja leksikaalisissa yhteisesiintymissä sekä sanojen semanttisessa preferenssissä ja semanttisessa prosodiassa**. Viimeiset neljä alaryhmää koostuvat puolestaan tapauksista, jotka lähentelevät jo virheitä siinä mielessä, että ne eivät ole enää välttämättä muodollisesti, semanttisesti tai kieliopillisesti hyväksyttäviä nyky-yleiskielen normien puitteissa. Nämä alaryhmät ovat **uudismuodosteet, sanakirjan sudenkuopat ja transfer-tapaukset, virheelliset nominivalinnat sekä virheelliset verbinvalinnat**.

4.2.1. Yksittäiset lähimerkityksiset sanat

Yksittäisten lähimerkityksisten sanojen alaryhmään kuuluvat tapaukset, joissa virkkeen kontekstiin sopisi suomenoppijan valitseman sanan sijaan jokin toinen vaihtelevassa määrin lähimerkityksinen sana. Epätyypillisuus ei siis ilmene myötäesiintymässä vaan yksittäisessä sanassa, jonka sopimattomuus selviää koko virkkeen kontekstista. Kutsun **sanapariksi** sitä vaihtoehtojen paria, joka koostuu suomenoppijan epätyypillisestä sananvalinnasta ja sen oletettavasti tyypillisemmästä vastineesta.

Kuuluvaisen (2015: 33) aiemman havainnon tavoin lähimerkityksisyys muodostaa tässäkin tutkimuksessa jatkumon, jonka yhdessä ääripäässä on lähes synonyymisia sanapareja ja toisessa taas niitä, joiden vaihtoehdot ovat jo semanttisesti kauempana toisistaan. Pääosa Kuuluvaisen (mts. 34) tapauksista on merkitykseltään kaukaisempia sanapareja, mutta tämä ei näytä enää pätevän edistyneellä tasolla, sillä suurin osa aineistoni sanapareista koostuu lähimerkityksisistä vaihtoehdoista. Tämän ryhmän epätyypillisyydet jakautuvat viiteen keskittymään, jotka ovat kontekstiin sopimattomat substantiivit, adjektiivit, kvanttorisanat, yksittäiset verbit sekä modaaliverbit.

Kontekstiin sopimattomat substantiivit. Useimmiten esiintyvä epätyypillinen substantiivinvalinta tässä keskittymässä on sana *aate*. Sitä käytetään seuraavissa esimerkeissä merkityksessä 'ajatus, päähänpisto, päähänpälkähdys, mielenjohtuma, idea', mikä onkin yksi sen Kielitoimiston sanakirjan antamista merkityksistä (KS s.v. *aate*). AATE-sanankondanssihaun yli 1 550 osumasta käy kuitenkin ilmi, että sanaa käytetään ylivoimaisesti merkityksessä 'yksilön maailmankatsomusta hallitseva t. yhteisölle ominainen yleisluonteinen näkemys, käsitys t. pyrkimys' (mp.). Tyypillisempi vaihtoehto olisi siis *idea* esimerkissä (1), *aihe* esimerkissä (2) ja *ajatus* esimerkissä (3).

- (1) *Onnistuin lopettamaan romaanin, mikä ei ikinä käynyt sen toisen osan tapauksessa, sillä minulta lieenee puuttunut **aatteita** (pro **ideoita**) [romaanin kirjoittamisesta]. (PU0005i)*
- (2) *Toinen **aate** (pro **aihe**), josta keskustelijat puhuvat, on eduskunnan edustajien määrän vähentäminen. (PU0006)*
- (3) *Olen kirjekurssin jälkeen unohtanut niin paljon, että **aate** (pro **ajatus**) yliopistossa opiskelemisesta pelotti alusa... (PU0022)*

Toinen lähimerkityksinen sanapari esiintyy esimerkissä (4), jossa sanaa *tunnelma* on käytetty viittaamaan työympäristön olosuhteisiin. Hakusanojen TYÖTUNNELMA ja TYÖILMAPIIRI kondanssitulokset kuitenkin paljastavat, että *ilmapiiri*-sanaa suositetaan ylivoimaisesti työhön liittyvissä konteksteissa, sillä osumia hakusanalla TYÖTUNNELMA oli vain kaksi, kun taas hakusana TYÖILMAPIIRI antoi lähes 770 osumaa.

- (4) *Tutkimukset osoittavat, että **tunnelmat** (pro **ilmapiirit**) [työympäristössä] ovat tärkeitä. (TS0009g)*

Ryhmässä esiintyy myös esimerkkien (5) ja (6) kaltaisia semanttisesti kaukaisempia sanapareja, joissa suomenoppijan käyttämä sana muistuttaa muodoltaan jossain määrin olettavasti tyypillisempää vaihtoehtoa. Vaikka vaihtoehtoja ei voida enää pitää synonyymisina, myös niiden välillä on nähtävissä jonkinlainen semanttinen yhteys.

- (5) *Hyve (pro hyvä puoli) siinä olisi, että meillä olisi todella erikoinen vaakuna ja varmasti jokainen ulkomaalainen tiesi sen.* (PU0003h)
- (6) *Heidät murhattiin erittäin julmasti, joitakin ruumiinosia löytyy jopa lähimetsästä, eikä mitään tekosyytä (pro syytä ~ selitystä) näytä olevan.* (PU0022d)

Kontekstiin sopimattomat adjektiivit. Yksittäisiä epätyypillisiä adjektiivivalintoja ei ole aineistossani paljon, sillä substantiivin määritteenä adjektiivit yleensä muodostavat luonnostaan vähintäänkin kahden sanan myötäesiintymiä. Tapauksia on yhteensä vain kaksi, ja molemmissa adjektiivi esiintyy lauseen predikatiivina. Esimerkissä (7) suomenoppija ei ole välttämättä tiennyt, että paremmuusjärjestyksessä ensimmäisellä on oma sanansa (KS s.v. *ykkönen*). Myös esimerkin (8) vaihtoehdot ovat merkityksiltään hyvin lähellä toisiaan, mutta kun virkkeen konteksti viittaa merkitykseen 'ehdottoman tarpeellinen, ehdoton', vain *välttämätön* sopii siihen (KS s.v. *väistämätön, välttämätön*).

- (7) *Yhtiön päätuotteita kuten maitojuoma, puhdistettu vesi ja sekoitettu puuro pidetään ensimmäisenä (pro ykkösenä) Kiinassa monta vuotta, varsinkin maitojuoman ja pulloveden tulot voivat verrata suureihin yrityksiin maailmassa.* (KI0016s)
- (8) *Ja kysyminen ei oikeasti riitä: on väistämätöntä (pro välttämätöntä), että kehitysmaat olisivat itse aktiivisia omien maiden tulevaisuuden luomisessa.* (SL0005h)

Kontekstiin sopimattomat kvanttorisanat. Tämän keskittymän tapaukset koskevat lähinnä kvanttoriadverbin *paljon* epätyypillinen käyttö. Kyseinen kvanttoriadverbi ilmaisee muun muassa substantiivilausekkeen tarkoittaman joukon tai aineen määrää sekä yhteydestä ilmenevän objektitarkoitteen tai tekemisen määrää (VISK § 657). Seuraavassa kolmessa esimerkissä *paljon*-sanaa on käytetty sille sopimattomassa lauseenjäsenen roolissa: intransitiivilauseen subjektin määritteenä esimerkeissä (9) ja (10) ja nollapersoonalauseen adverbialina esimerkissä (11). Mikään näistä rooleista ei kuulu *paljon*-sanana tehtäviin (ks. VISK § 657), ja tyypillisempi vaihtoehto kaikissa tapauksissa olisi suhteellista

lukumäärää ilmaiseva kvanttoripronomini *moni* (ks. VISK § 764), joka toimii lauseessa sellaisenaan tai adjektiivin tavoin substantiivin määritteenä.

- (9) *Paljon asioita* (pro *moni asia*) riippuu vain omista valinnoistaan, mutta kysymys on, voiko yksilö muuttaa maailmassa mitään? (TS0011g)
- (10) *Kommunismien aikana paljon* (pro *monet*) suomalaiset siirtyivät Karjalan tasavaltaan. (UN0013b)
- (11) *Oppimisen aikana paljoilla* (pro *monilla*) voi tapahtua niin. (VI0399a)

Kontekstiin sopimattomat yksittäiset verbit. Myös epätyypillisten verbinvalintojen sanaparit koostuvat useimmiten lähimerkityksistä vaihtoehtoista, kuten esimerkin (12) *kohdata* ja *tavata*. Molemmat verbit sisältävät merkityksen 'tulla kosketuksiin', mutta haksanojen TAVATA ja KOHDATA konkordanssitulokset paljastavat hienoisen merkityseron. Niiden perusteella verbi *tavata* näyttää olevan yleisemmin käytössä, kun puhutaan halutuista tai suunnitelluista tapahtumista, kun taas verbiä *kohdata* käytetään yleisemmin viittaamaan satunnaisiin tai jopa epätoivottuihin tapahtumiin. Lisäksi Kielitoimiston sanakirja antaa merkityksen 'tulla kosketukseen vars. seurustelu-, keskustelu- tms. mielessä tois(t)en kanssa' vain *tavata*-verbille (KS s.v. *tavata*, *kohdata*).

- (12) *Hän asuu Lontoossa ja siksi emme voi kohdata* (pro *tavata*) usein. (PU0026)

Myös esimerkin (13) sanaparin vaihtoehdot ovat lähimerkityksisiä. Sekä *vaihtua* että *muuttua* tarkoittavat, että yhdestä asiasta tulee toisenlainen (KS s.v. *vaihtua*, *muuttua*), ja kumpikin sana sopii luontevasti virkkeeseen niin kielipollisesti kuin semanttisestikin. Kuitenkin tässä suomenoppijalta on jäänyt huomaamatta se merkitysero, että kun kieli *vaihtuu*, se korvataan kokonaan uudella kielellä, mutta kun se *muuttuu*, sama kieli kehittyy toisenlaiseksi ajan myötä.

- (13) *Tämä aihe on kovin mielenkiintoinen ja ajankohtainen, koska kieli vaihtuu* (pro *muuttuu*) jatkuvasti, siis slangi *vaihtuu* (pro *muuttuu*) myös. (UN0013b)

Esimerkin (14) verbit alkavat jo olla semanttisesti kauempana toisistaan, vaikka ne voidaan kääntää samalla sanalla muun muassa englanniksi (*stop*). Kun tässä viitataan aiheen kirjoittamiseen loppuun, verbi *pysähtyä* ei sovi kontekstiin, koska siltä puuttuu

lopettaa-verbin merkitys 'tehdä jk valmiiksi, saattaa jk loppuun' (KS s.v. *pysähtyä*, *lopettaa*).

- (14) *(Anteeksi, minä tiedän että juttuni on liian pitkä, mutta se sujuu niin hyvin, että minä en pysty pysähtymään (pro lopettamaan).)* (HO0001k)

Kontekstiin sopimattomat modaaliverbit. Huomionarvoista on myös se, että erityisesti modaaliverbit näyttävät aiheuttavan edistyneillekin suomenoppijoille vaikeuksia. **Modaalisuus** on semanttinen alue, joka koskee asiointilan todenmukaisuutta ja toteutumismahdollisuuksia (VISK § 1551). Aineistoni epätyypillisyydet liittyvät pääosin asiointilan mahdollisuutta ja luvallisuutta ilmaisevien modaaliverbien valintaan. Yleisimpiä ovat tapaukset, joissa yleiskäyttöinen *voida*-verbi sekoitetaan toisen modaaliverbin kanssa. Sekoittaminen johtunee siitä, että vaikka *voida*-verbiä on mahdollista käyttää dynaamisen, deonttisen tai episteemisen mahdollisuuden ilmaisemiseen, se ei ole aina tulkittavissa yhdeksikään niistä modaalisuuden lajeista (VISK § 1566). Erilaisten modaaliverbien vivahteisiin perehtymätön suomenoppija saattaa näin päätyä epätyypilliseen vaihtoehtoon sopivaa modaaliverbiä valitessaan.

Yleiskäyttöisen *voida*-verbin ja deonttisen *saada*-verbin sekoittaminen. Modaaliverbi *saada* edustaa deonttista modaalisuutta, joka koskee muun muassa puheyhteisön normeja ilmaisevaa lupaa (VISK § 1551). Vaikka deonttista modaalisuutta voidaan joissakin tilanteissa ilmaista *voida*-verbilläkin, yksiselitteisempi modaaliverbi kuten *saada* tarvitaan, kun halutaan selvyiden vuoksi välittää sallittuuden merkitystä itse verbissä (VISK § 1569). Modaaliverbin valinta esimerkissä (15) antaa ymmärtää, että alivaltiosihteeri ei kykene kuulemaan kissasta, vaikka tarkoitus olikin luultavasti ilmaista, että hänen ei pitäisi antaa kuulla siitä.

- (15) *Mutta heidän pitää olla varovaisia, koska alivaltiosihteeri ei voi (pro saa) kuulla, että siellä on kissa.* (PU0023n)

Esimerkeissä (16) ja (17) *voida*-verbin käyttöä pidettäisiin luultavasti tyyppillisenä, jos virkkeet olisivat preesensissä. Verbin kolligaatioon kuuluu kuitenkin se, että myönteisessä lauseessa se ei juurikaan esiinny aktiivin imperfektin muodossa. Tämä näkyy hyvin selvästi hakusanan VOIDA (AKT IMPF) konkordanssituloksista: vaikka osumia olikin

runsaat 1 010, tulosten kvalitatiivinen tarkastelu paljastaa, että suurin osa ei olekaan oikeastaan imperfektissä.

- (16) *Vuodesta 1859 alkaen suomalaiset voivat (pro **saivat**) viedä Venäjälle tavaroitaan ilman tullia.* (PU0007c)

- (17) *Sodan jälkeen Jack Starks lopulta **voi** (pro **sai**) mennä kotiin.* (PU0027c)

Myös verbiä *saada* käytetään silloin, kun sopivampi vaihtoehto olisi *voida*. Esimerkeissä (18) ja (19) on käytetty sallittuutta ilmaiseva *saada*-verbiä, kun kyse on pikemminkin jonkin asian toteutumismahdollisuudesta. Tällöin yleiskäyttöinen *voida*-verbi sopisi paremmin lauseiden kontekstiin, erityisesti jos esimerkki (19) muutetaan preesensmuotoon.

- (18) *Toisaalta, tiedän ettei aina **saa** (pro **voi**) vaikuttaa onnellisuuteen.* (PU0023h)

- (19) *Korpus tarkoitti osakielen tai murteen tekstien kokoelmaa, josta tiettyjä analyysimenetelmiä soveltaessa **sai** (pro **voi**) kielentutkija johtaa sen kielen kieliooppia.* (VI0393b)

Yleiskäyttöisen *voida*-verbin ja dynaamisten voimisverbien sekoittaminen. Dynaaminen modaalisuus tarkoittaa ”fyysistä mahdollisuutta ja kykyä tai välttämättömyyttä ja pakkoa toimia” (VISK § 1554). Spesifisillä dynaamista mahdollisuutta merkitsevillä verbeillä ilmaistaan muun muassa kykenemistä, joka voi olla joko fyysinen (esim. *pystyä*, *osata*) tai psyykkinen (esim. *malittaa*) (VISK § 1567). Esimerkissä (20) on käytetty fyysistä kykenemistä ilmaisevaa *pystyä*-verbiä, vaikka kyse on pikemminkin psyykkisestä voimisesta. Tällöin yleistulkintainen *voida*-verbi sopisi kontekstiin paremmin.

- (20) *Vaikka on itsestään selvää, että ryöstäjän teko oli pelkkää paha, joka aiheutti sekä tekijälle, että monille muille ihmisille ongelmia ja surua, häntä ei **pysty kutsumaan** (pro **voi kutsua**) huonoksi ihmiseksi.* (PU00221)

Esimerkissä (21) on puolestaan käytetty yleistulkintaista *voida*-verbiä – myös epätyylillisesti aktiivin imperfektin muodossa –, kun kyseessä on selvästi fyysinen kykeneminen jatkamaan matkaa ilman, että vuori olisi edessä esteenä.

- (21) *Hän käänsi vuoren ja veljet **voivat jatkaa** (pro **pystyivät jatkamaan**) matkaa.* (PU0016g)

Myös osaamista ilmaisevaa modaaliverbiä sekoitetaan yleiskäyttöisen *voida*-verbin kanssa. Kun esimerkissä (22) puhutaan kielitaidosta, kontekstiin sopisi paremmin spesifimpi *osata*-verbi kuin yleiskäyttöinen *voida*. Myös hakuyhdistelmän OSATA & PUHUA konkordanssituloksissa esiintyy huomattavasti enemmän kieleen viittaavia käyttöyhteyksiä kuin hakuyhdistelmän VOIDA & PUHUA konkordanssituloksissa.

- (22) *Nuoret puolalaiset usein osaavat sitä [englantia] ainakin vähän, mutta jo keski-ikäiset tuskin **voivat** (pro **osaavat**) sanoa mitään.* (PU0005b)

Esimerkissä (23) *osata*-verbiä on käytetty epätyypillisesti predikaattina lauseessa, jonka subjekti ei ole elollinen tarkoite. Verbin merkitykseen liittyy tietoisuus ja aistivuus (KS s.v. *osata*), joka puuttuu kulttuuriperinteen kaltaiselta abstraktilta käsitteeltä. Verbin käyttöpreferenssejä vahvistaa hakusanan OSATA (AKT) konkordanssitulokset, joiden mukaan verbiä käytetään ylivoimaisesti elollisten subjektien yhteydessä.

- (23) *Ei mikään kulttuuriperinne **osaa tappaa** (pro **voi tappaa**) toista kulttuuriperinnettä.* (TS0014a)

Fyysistä voimista ilmaisevien modaaliverbien lisäksi aineistossani esiintyy myös ainakin yksi tapaus, jossa yleiskäyttöistä verbiä *voida* sekoitetaan psyykkistä voimista ilmaisevan *maltaa*-verbin kanssa. Esimerkistä (24) välittyy merkitys, että kirjoittaja ei kykene odottamaan lumen tuloa, vaikka tarkoitus olikin luultavasti ilmaista, että hän odottaa sitä innolla. Saman merkityseron näkee myös hakuyhdistelmien EI + VOIDA + ODOTTAA ja EI + MALTTAA + ODOTTAA konkordanssituloksista.

- (24) *En **voi** (pro **malta**) odotta, kun aloittaa sataa lunta!* (PU00026)

Episteemisten modaaliverbien sekoittaminen. Episteeminen modaalisuus liittyy ”puhujan uskoon, tietoon tai päätelmiin asiointilan mahdollisuudesta, todennäköisyydestä tai varmuudesta” (VISK § 1551, § 1556). Esimerkin (25) sanaparissa sekä *taittaa* että *mahtaa* ilmaisevat mahdollisuutta, ja kumpikin tulkitaan yksinomaan episteemisesti (VISK § 1562). Vaikka verbit ovatkin lähimerkityksisiä, verbin *taittaa* käyttökonteksti liittyy jonkin asian ilmaisemiseen todennäköiseksi, luultavaksi tai mahdolliseksi, kun taas verbin *mahtaa* käyttökonteksti koskee pikemminkin arvelun ilmaisemista (KS s.v. *taittaa*,

mahtaa). Koska esimerkissä puhutaan vanhempien juttelun sisällöstä eikä tapahtuman mahdollisuudesta, verbi *mahtaa* on siihen sopivampi vaihtoehto.

- (25) *Aluksi hän ei saanut selväksi melkein mitään, ei varsinkaan sitä mistä hänen vanhemmat taitaa (pro mahtavat) jutella niin huolistuneella äänellä, kuten hän on huomannut saavuttuaan oven viereen. (UN0004)*

4.2.2. Intensiiteettipartikkelit

Intensiiteettipartikkelit luonnehtivat muun muassa adjektiivin tai adverbien ilmaiseman laadun tai määrän astetta (VISK § 664, § 853). Osa ryhmän tapauksista koskee intensiiteettipartikkelien epätyypillistä kollokaatiota ja semanttista preferenssiä, jolloin pitkälti synonyymisia mutta tyyliarvoltaan toisistaan eroavia intensiiteettipartikkeleita sekoitetaan keskenään. Osa liittyy taas intensiiteettipartikkelien kolligaatioon, jolloin niitä käytetään niille epätyypillisissä kieliopillisissa yhteyksissä.

Intensiiteettipartikkelien alaryhmä jakautuu kuuteen keskittymään. Niistä ensimmäiset kaksi ovat intensiiteettipartikkelien *aivan* ja *oikein* kollokaatio ja semanttinen preferenssi. Kolmas keskittymä koskee puolestaan intensiiteettipartikkelin *sangen* kollokaatiota, ja neljäs liittyy intensiiteettipartikkelin *kovin* kolligaatioon ja semanttiseen preferenssiin. Viimeiset kaksi keskittymää taas ovat intensiiteettipartikkelien *hyvin* ja *erittäin* kolligaatio.

Intensiiteettipartikkelin *aivan* kollokaatio ja semanttinen preferenssi. Esimerkeissä (26) ja (27) sanaa *aivan* on käytetty negatiivisen merkityksen sisältävän adjektiivin määrittäenä. Hakuyhdistelmän AIVAN + ADJ konkordanssitulokset kuitenkin paljastavat, että sanan määrittämät adjektiivit ovat suurimmilta osin positiivisia tai neutraaleja. Hakutulosten yleisimmät kymmenen yhdistelmää ovat AIVAN + UUSI, ERILAINEN, MUU, OMA, MAHTAVA, TAVALLINEN, USKOMATON, TOISENLAINEN, UUDENLAINEN sekä LOISTAVA, ja vaikka tuloksissa esiintyykin joitain negatiivisia vierusadjektiiveja kuten KÄSITTÄMÄTÖN, TURHA, VÄÄRÄ ja HIRVEÄ, ne ovat silti suhteellisen harvinaisia.

- (26) *On aivan (pro hyvin ~ todella ~ erittäin) surullista se, että hänellä ei ole lämpimiä tunteita perheensä kohtaan eikä hän itse tunne yhteystä heidän kanssa. (PU0004i)*

- (27) *Hän ei osallistu tapahtumiin, mutta ei hän ole **aivan** (pro **täysin**) välinpitämätön päähenkilöitä kohtaan.* (VE0129a)

Aivan kaikkien positiivisten adjektiivien kanssa sana ei kuitenkaan näytä kollokoivan. Esimerkissä (28) sitä on käytetty määrittämään adjektiivia *mielenkiintoinen*, mikä ei liene kovin tyypillistä, sillä hakuyhdistelmä AIVAN + MIELENKIINTOINEN antoi yhden ainoan osuman. Sen sijaan osumia oli yli 380 hakuyhdistelmällä HYVIN + MIELENKIINTOINEN ja 360 hakuyhdistelmällä TODELLA + MIELENKIINTOINEN, joten kumpikin vaihtoehto sopisi *mielenkiintoinen*-adjektiivin määritteeksi paremmin kuin *aivan*.

- (28) *Helmikuun kahdennenkymmenennenseitsemännen päivän ”Knalli ja sateenvarjo”-kuunnelman jakso on **aivan** (pro **hyvin** ~ **todella**) mielenkiintoinen, sillä pätkässä on sanaleikkejä ja siinä tapahtuu kaikenlaisia väärinkäsityksiä, jotka aiheuttavat koomisuutta.* (PU00311)

Samoin *aivan* ei näytä kollokoivan esimerkin (29) *tärkeä*-adjektiivin kanssa, sillä hakuyhdistelmä AIVAN + TÄRKEÄ antoi yhden ainoan osuman, kun taas osumia oli 1 830 hakuyhdistelmällä HYVIN + TÄRKEÄ, lähes 1 870 hakuyhdistelmällä TODELLA + TÄRKEÄ sekä yli 3 680 hakuyhdistelmällä ERITTÄIN + TÄRKEÄ.

- (29) *Itsenäisyys on **aivan** (pro **hyvin** ~ **erittäin** ~ **todella**) tärkeää.* (VE0093f)

Intensiteettipartikkelin *oikein* kollokaatio ja semanttinen preferenssi. Esimerkeissä (30) ja (31) sanaa *oikein* on käytetty negatiivisen merkityksen sisältävän adjektiivin määritteenä. Tämä on ristiriidassa esimerkiksi Jantusen (2004) tutkimuksen kanssa, jonka mukaan *oikein*-sanalla on vahva semanttinen preferenssi 'arvoa' ja 'tunnetta' merkitseviin sanoihin. Tapausten epätyypillisyyden vahvistavat hakuyhdistelmän OIKEIN + ADJ konkordanssitulokset, joiden mukaan yleisimmät kymmenen yhdistelmää ovat OIKEIN + HYVÄ, MUU, MUKAVA, TYYTYVÄINEN, HIENO, PAHA, ISO, HUONO, KIVA sekä KOVA. Näistä adjektiiveista ainoastaan PAHA ja HUONO sisältävät yksinomaan negatiivisen merkityksen.

- (30) *Toisaalta oli **oikein** (pro **hyvin**) vaikeaa selittää yhteiskunnalle että Andrzej Lepper joka oli rikollinen, tulee varapääministeriksi.* (PU0024)
- (31) *Hän on **oikein** (pro **hyvin** ~ **erittäin** ~ **todella**) onneton.* (VE0100)

Samoin kuin *aivan* myöskään *oikein* ei näytä kollokoivan ihan kaikkien positiivisten adjektiivien kanssa. Sanan käyttö *tärkeä*-adjektiivin määritteenä esimerkeissä (32) ja (33) ei liene tyypillistä, sillä hakuyhdistelmä OIKEIN + TÄRKEÄ antoi alle kymmenen osumaa. Sen sijaan osumia oli runsaasti hakuyhdistelmillä HYVIN + TÄRKEÄ, TODELLA + TÄRKEÄ sekä ERITTÄIN + TÄRKEÄ, kuten *aivan*-sanana osalta kävi jo ilmi.

- (32) *Toisaaltakin suomalaisten historioitsijoiden mukaan Juhanan avioliitto puolalaisen infantin kanssa ja heidän oleskelunsa Suomessa oli tämän maan historian **oikein** (pro **hyvin** ~ **erittäin** ~ **todella**) tärkeä tapahtuma.* (PU0001e)
- (33) *Minusta tuntuu, että tämä kohta on **oikein** (pro **hyvin** ~ **erittäin** ~ **todella**) tärkeä Sinuhen elämässä, nyt hän tietää, kenen poika hän on ja löytää vastaukset kysymyksiinsä.* (VE0100)

Esimerkeissä (34) ja (35) sanaa *oikein* on käytetty kieltolauseissa, joihin *kovin* sopisi paremmin. Hakuyhdistelmä EI + OLLA + OIKEIN + VAHVA antoi yhden ainoan osuman, kun taas niitä oli yli 30 hakuyhdistelmällä EI + OLLA + KOVIN + VAHVA. Myöskään hakuyhdistelmä EI + VER + OIKEIN + USEIN ei antanut yhtään osumaa, kun taas niitä oli 90 hakuyhdistelmällä EI + VER + KOVIN + USEIN.

- (34) *Se [punainen kaartti] ei ollut sentään **oikein** (pro **kovin**) vahva, sillä sitä heikenti sisällisiä riitoja ja johdon päättämättömyys.* (PU0005d)
- (35) *Osallistuin siihen mielelläni, vaikka piti olla yliopistossa monta tuntia aikaisemmin, koska sentapaisia tapahtumia ei järjestetä **oikein** (pro **kovin**) usein ja siksi täytyy nauttia jokaisesta tilaisuudesta.* (PU0022)

Intensiteettipartikkelin *sangen* kollokaatio. Esimerkissä (36) sanaa *sangen* on käytetty *muu*-adjektiivin määritteenä. Kyseiset sanat eivät kuitenkaan näytä kollokoivan toistensa kanssa, sillä konkordanssihaku yhdistelmällä SANGEN + MUU ei antanut yhtään osumaa. Sen sijaan hakuyhdistelmä AIVAN + MUU antoi yli 900 osumaa, eli *aivan* sopisi *muu*-adjektiivin määritteeksi paljon paremmin kuin *sangen*.

- (36) *Kontrastiivisen analyysin (CA – contrastive analysis) tuloksien tarkastelu korpuspohjaisesti antaa enemmän mahdollisuuksia poikkeuksien selittämiseen: onko kyseessä lähdekielen ja kohdekielen välinen siirtovaikutus, universaali oppijankielen piirre tai jotain **sangen** (pro **aivan**) muuta.* (VI0398a)

Intensiteettipartikkelin *kovin* kolligaatio ja semanttinen preferenssi. Aineistossani esiintyy useita tapauksia, joissa sanaa *kovin* on käytetty adjektiivin määritteenä myönteisissä lauseissa. Vaikka tällainen käyttö on yleistä natiivisuomessakin, *kovin* esiintyy kielteisissä lauseissa huomattavasti useammin: esimerkiksi konkordanssihaku yhdistelmällä *olen, olet, on, olemme, olette, ovat, ollaan* + [0–1 sana] + KOVIN + ADJ antoi lähes 1 270 osumaa, kun taas osumia hakuyhdistelmällä *ole* + [0–1 sana] + KOVIN + ADJ oli peräti lähes 3 960.

Sanan *kovin* käyttö myönteisissä lauseissa ei siis ole sinänsä epätyypillistä. Sen käyttöyhteyksien tarkastelu myönteisissä lauseissa kuitenkin paljastaa, että yleisimmät kymmenen yhdistelmää ovat OLLA + KOVIN + ERILAINEN, PIENI, VAIKEA, SUURI, MAHDOLLINEN, LYHYT, TUTTU, RAJALLINEN, YLEINEN ja HELPPÖ. Toisin sanoen *kovin* esiintyy myönteisissä lauseissa useimmiten neutraalien tai negatiivisten adjektiivien yhteydessä.

Esimerkeissä (37) ja (38) *kovin* määrittää positiivista adjektiivia, mikä saattaa olla konkordanssitulosten perusteella epätyypillistä. Tämä pitää paikkansa esimerkin (37) osalta, sillä hakuyhdistelmä *olen, olet, on, olemme, olette, ovat, ollaan* + [0–1 sana] + KOVIN + TÄRKEÄ antoi alle kymmenen osumaa, kun taas hakuyhdistelmä *olen, olet, on, olemme, olette, ovat, ollaan* + [0–1 sana] + HYVIN + TÄRKEÄ antoi yli 1 020 osumaa. Esimerkki (38) saattaa puolestaan olla rajatapaus, sillä hakuyhdistelmän OLLA + [0–1 sana] + KOVIN + MIELENKIINTOINEN 15 osumasta neljä eli lähes kolmasosa liittyy *kovin*-sanon käyttöön myönteisessä lauseessa. Vaikka osumia hakuyhdistelmällä OLLA + [0–1 sana] + HYVIN + MIELENKIINTOINEN olikin huomattavasti enemmän (vajaat 230, joista kaikki lauseet olivat myönteisiä), se, voidaanko *kovin*-sanon käyttöä esimerkissä (38) pitää epätyypillisenä, jää epäselväksi.

(37) *Aleksanterin II:n aika oli **kovin** (pro **hyvin**) tärkeä ja edullinen Suomelle ja autonomian kehitykselle.* (PU0007c)

(38) *Se on **kovin** (pro **hyvin**) mielenkiintoinen minun mielestäni, että elokuvissa Suomessa mustalaiset tavallisesti puhuvat kaakkoismurteita.* (UN0013b)

Sanan *kovin* käyttö esimerkkien (39) ja (40) myönteisissä lauseissa kuulostanee puolestaan luontevammalta kuin edelliset esimerkit, sillä niissä sanan määrittämät adjektiivit

ovat merkitykseltään negatiivisia. Kuitenkin molemmissa tapauksissa *hyvin* on jälleen paljon yleisempi määrite myönteisessä lauseessa kuin *kovin*.

- (39) *Jos eläin on **kovin** pahassa kunnossa saadaan sen eläinsuojelulain mukaan lopettaa.* (PU0001j)
- (40) *Sir Henry on alussa **kovin** vihainen, mutta nähtyään heidät taas ulokkeella hän muuttuu taas herttäiseksi ja kutsuu miehet huoneeseen keskustelemaan.* (PU0002j)

Konkordanssitulosten perusteella voitaneen siis päättää, että sanaa *kovin* käytetään jossain määrin adjektiivin määritteenä myönteisissäkin lauseissa, joskaan se ei ole silloin yleensä tyypillisin vaihtoehto. Silloin kun sitä käytetään tällaisessa yhteydessä, se määrittää useammin neutraalia tai negatiivista adjektiivia ja harvemmin positiivista adjektiivia. Lisäksi sanan sopivuus myönteiseen lauseeseen riippuu paljolti kontekstista. Sanan vivahteisiin perehtymättömän suomenoppijan olisi siis syytä harkita sen käyttöä erityisesti positiivisissa yhteyksissä.

Intensiteettipartikkelin *hyvin* kolligaatio. Esimerkeissä (41) ja (42) sanaa *hyvin* on käytetty adjektiivin määritteenä kieltoverbin yhteydessä. Tällainen käyttö lienee kuitenkin epätyypillistä, sillä konkordanssihaku yhdistelmällä *ole* + [0–1 sana] + HYVIN + ADJ antoi vain hieman yli 30 osumaa, kun taas hakuyhdistelmällä *olen, olet, on, olemme, olette, ovat, ollaan* + [0–1 sana] + HYVIN + ADJ osumia oli melkein 23 800. Konkordanssitulosten perusteella voitaneen sanoa, että vaikka *hyvin*-sanon käyttö kieltoverbin yhteydessä sopii tiettyihin konteksteihin, sanaa käytetään ylivoimaisesti myönteisissä lauseissa.

- (41) *Joka kaupunki on täynnä kommunistisia rakennuksia, jotka eivät tavallisesti ole **hyvin** (pro **kovin**) tyylikkäitä.* (PU0005b)
- (42) *Minusta juuri tämä rakkausjuoni ei ole **hyvin** (pro **kovin**) onnistunut ja se tuntuu vähän keinotekoiselta ja teennäiseltä.* (PU0023l)

Intensiteettipartikkelin *erittäin* kolligaatio. Myös *erittäin* näyttää esiintyvän adjektiivin määritteenä ylivoimaisesti myönteisissä lauseissa, sillä konkordanssihaku yhdistelmällä *olen, olet, on, olemme, olette, ovat, ollaan* + [0–1 sana] + ERITTÄIN + ADJ antoi lähes 17 490 osumaa, kun taas hakuyhdistelmällä *ole* + [0–1 sana] + ERITTÄIN + ADJ osumia oli vain vajaat 40. Sanan käyttöä esimerkeissä (43) ja (44) voitaneen siis pitää

epätyypillisenä, sillä *erittäin* esiintyy silloin myönteisessä lauseessa. Sopivampi vaihtoehto kummassakin tapauksessa olisi *kovin*, joka esiintyy paljon useammin kieltoverbin yhteydessä.

- (43) *Vaikka hän ei ollut **erittäin** (pro **kovin**) taitava kirjailija, hänen teosten arvo ja merkitys ovat kistämättömiä.* (PU0022a)
- (44) *Elokuva ei kuitenkaan kertoo mitään **erittäin** (pro **kovin**) uutta tarinaa.* (SA0092c)

4.2.3. Muut kollokaatiot ja leksikaaliset myötäesiintymät

Muiden kollokaatioiden ja leksikaalisten myötäesiintymien alaryhmä koostuu tapauksista, joissa suomenoppija on tavoitellut jotain vakiintunutta ilmaisua tai kollokaatiota siinä kuitenkaan onnistumatta. Ryhmän epätyypilliset myötäesiintymät koostuvat pääosin kahdesta vierekkäisestä tai toistensa lähettyvillä olevasta sanasta. Viitataan kollokaatiolla koko käsiteltävään myötäesiintymään ja kollokaatilla siihen suomenoppijan käyttämään yksittäiseen sanaan, joka on vaihdettavissa sen tyypillisempään vaihtoehtoon. Yksittäisten lähimerkityksisten sanojen (ks. luku 4.2.1.) tavoin sanaparien vaihtoehdot ovat useimmiten semanttisesti lähellä toisiaan. Ryhmä jakautuu neljään keskittymään, jotka ovat epätyypilliset substantiivikollokaatit, adjektiivikollokaatit, adverbikollokaatit sekä verbi-kollokaatit.

Epätyypilliset substantiivikollokaatit. Tämä keskittymä koostuu muun muassa tapauksista, joissa verbin yhteyteen on valittu epätyypillinen substantiivi. Sanaa *seikka* on käytetty *järjestää*-verbin objektina esimerkissä (45) ja *palata*-verbin adverbialina esimerkissä (46), vaikka kummassakin tapauksessa sopivampi vaihtoehto olisi *asia*. Päätelmän vahvistavat hakuyhdistelmien JÄRJESTÄÄ & SEIKKA (OBJ) ja JÄRJESTÄÄ & ASIA (OBJ) konkordanssitulokset: edellinen hakuyhdistelmä ei antanut yhtään kollokaatio-osumaa, kun taas kollokaatio-osumat muodostivat merkittävän osan jälkimmäisen hakuyhdistelmän vajaasta 640 osumasta. Samoin konkordanssihaku yhdistelmällä PALATA & SEIKKA (ILL) antoi yhden ainoan osuman, kun taas kollokaatio-osumia hakuyhdistelmällä PALATA & ASIA (ILL) oli peräti 680.

- (45) *Suomen herttuana ja käskyhaltijana, Juhana tuli Turkuun, missä hän heti alkoi järjestää herttuakuntansa **seikkoja** (pro **asioita**). (PU0001e)*
- (46) *Sitten hän palaa julisteseikkaan (pro julisteasiaan) ja käskee kaiken mennä kotiin, koska kello on jo paljon. (PU0002j)*

Esimerkissä (47) on käytetty ilmausta *kysyä jkn mieltä*, vaikka tyypillisempi vaihtoehto olisi *kysyä jkn mielipidettä*. Konkordanssihakua yhdistelmällä KYSYÄ & MIELI (OBJ) antoi 110 osumaa, joista suurin osa oli kuitenkin muodossa KYSYÄ – – *mitä mieltä* eikä siis etsimiäni kollokaatio-osumia. Hakuyhdistelmä KYSYÄ & MIELIPIDE (OBJ) antoi puolestaan yli 1 030 osumaa, josta suurin osa on nimenomaan kollokaatio-osumia. Konkordanssituloksista siis ilmenee, että *kysyä jkn mielipidettä* on paljon tyypillisempi ilmaus kuin *kysyä jkn mieltä*.

- (47) *Hän sanoo kansanedustajien palkkioiden noustun, ehdottaa, että suomalaisten parlamentaarikkojen määrä vähennettäisiin 150:en ja kysyy muiden **mieltä** (pro **mielipidettä**). (PU0005)*

Epätyypilliset adjektiivikollokaatit. Tämä keskittymä koostuu tapauksista, joissa substantiivin määritteeksi on valittu epätyypillinen adjektiivi. Esimerkissä (48) *sää* on kuvattu *kannustavaksi*, vaikka tyypillisempi adjektiivi olisi *suotuista*. Tapauksen epätyypillisuus varmistuu konkordanssituloksista, joissa hakuyhdistelmä SÄÄ & KANNUSTAVA ei antanut yhtään osumaa, kun taas hakuyhdistelmän SÄÄ & SUOTUISA reilusta 130 osumasta suurin osa oli juuri kollokaatio-osumia.

- (48) *Vaikka talvisää ei ollut **kannustava** (pro **suotuista**), he tulivat seisomaan. (PU0002b)*

Esimerkissä (49) adjektiivia *vahva* on käytetty *pula*-sanana määritteenä, vaikka tyypillisempi vaihtoehto olisi *kova*. Tämä näkyy hyvin selvästi konkordanssituloksissa, joissa hakuyhdistelmä VAHVA & PULA ei antanut yhtään kollokaatio-osumaa, mutta suurin osa hakuyhdistelmän KOVA & PULA lähes 370 osumasta oli juuri kollokaatio-osumia.

- (49) *Venäjänkielisten työntekijöiden pula on **vahva** (pro **kova**) Kaakkois-Suomessa, jossa käy paljon venäläisiä turisteja. (VE0130e)*

Epätyypilliset adverbikollokaatit. Tämä keskittymä koostuu tapauksista, joissa epätyypillinen adverbi on valittu jonkin sanan määritteeksi. Esimerkissä (50) adverbia *väärin* on käytetty *sujuu*-verbin määritteenä. Hakuyhdistelmän SUJUA & VÄÄRIN konkordanssitulokset kuitenkin paljastavat, että kyseiset sanat eivät kollokoi toistensa kanssa, sillä yhtään kollokaatio-osumaa ei ollut. Sen sijaan hakuyhdistelmä SUJUA & HUONOSTI antoi lähes 130 osumaa, joista suurin osa on juuri kollokaatio-osumia.

- (50) *Hän haluaa kostaa ryssälle, mutta sillä välin kaikki sujuu **väärin** (pro **huonosti**).* (HO0001f)

Esimerkissä (51) adverbia *vakavasti* on käytetty verbin *kritisoida* määritteenä. Kyseiset sanat eivät kuitenkaan näytä kollokoivan keskenään, sillä konkordanssihaku yhdistelmällä KRITISOIDA & VAKAVASTI ei antanut yhtään osumaa. Sen sijaan kollokaatio-osumia oli hakuyhdistelmällä KRITISOIDA & ANKARASTI vajaat 70.

- (51) *Alman surullisen kokemuksen taustana Minna Canth kritisoi **vakavasti** (pro **ankarasti**) keskiluokan miehen maailmaa 1800-luvulla.* (KI0012)

Epätyypilliset verbikollokaatit. Tämä keskittymä koostuu tapauksista, joissa epätyypillinen verbi on valittu jonkin sanan yhteyteen. Tällaisia epätyypillisiä verbinvalintoja on tehty muun muassa lauseen objektina esiintyvien substantiivien yhteydessä. Esimerkissä (52) verbiä *tehdä* on käytetty viittaamaan valokuvaukseen, vaikka tyypillisempi vaihtoehto olisi *ottaa*. Konkordanssihaku yhdistelmällä TEHDÄ & VALOKUVA (OBJ) antoi vajaat sata osumaa, joista vain kymmenkunta oli kollokaatio-osumia. Sen sijaan hakuyhdistelmän OTTAA & VALOKUVA (OBJ) lähes 1 040 osumasta suurin osa oli kollokaatio-osumia.

- (52) ***Teimme** (pro **otimme**) ehkä vähemmän valokuvia, painoimme niitä palattamme kotiin, katsoimme niitä yhden kerran ja sitten ne menivät laatikkoon.* (PU0001c)

Esimerkissä (53) verbiä *suurentaa* on käytetty *mahdollisuus*-objektin yhteydessä. Konkordanssihaku yhdistelmällä SUURENTAA & MAHDOLLISUUS (OBJ) antoi yhden ainoan kollokaatio-osuman, kun taas kollokaatio-osumat muodostivat suurimman osan hakuyhdistelmän PARANTAA & MAHDOLLISUUS (OBJ) yli 690 osumasta. Esimerkissä (54) verbiä *puhua* on käytetty *mielipide*-objektin yhteydessä, mikä lienee myös epätyypillistä. Tämä näkyy hakuyhdistelmän PUHUA & MIELIPIDE (OBJ) konkordanssituloksissa, sillä lähes 60

osumasta alle kymmenen oli kollokaatio-osumia. Sen sijaan suurin osa hakuyhdistelmien KERTOJA & MIELIPIDE (OBJ) ja SANOA & MIELIPIDE (OBJ) osumista oli kollokaatio-osumia.

- (53) *Näkemällä todellisuuden hyvät puolet ja katsomalla kaikkea huumorintajua käyttäen ihminen merkittävästi **suurentaa** (pro **parantaa**) mahdollisuutensa olla onnellinen.* (PU0027i)
- (54) *Suomalaisessa opiskelukulttuurissa hyvä on se asia, että opiskelijat voivat olla eri mieltä opettajan kanssa ja **puhua** (pro **kertoa** ~ **sanoa**) omia mielipiteitään.* (VE0110i)

Esimerkissä (55) verbi *kiljua* on valittu *leijona*-subjektin predikaatiksi. Ilman todistavaa korpusaineistoaakin lienee selvää, että leijonat eivät tyypillisesti *kilju* vaan ne *karjuvat*.

- (55) *Sillä [leijonalla] oli runsas harja ja se [leijona] **kiljui** (pro **karjui**) koko ajan kovalla äänellä.* (PU0032b)

Substantiivien lisäksi epätyypillisiä verbinvalintoja on tehty myös adverbiaalien yhteydessä. Esimerkki (56) on rajatapaus siinä mielessä, että ilmaus *lähteä eläkkeelle* on selvästi käytössä natiivisuomessakin, sillä konkordanssihaku yhdistelmällä LÄHTEÄ & ELÄKE (ALL) antoi yli 191 osumaa, josta suurin osa oli kollokaatio-osumia. Ilmauksen käyttötaajuus on kuitenkin pieni verrattuna hakuyhdistelmän JÄÄDÄ & ELÄKE (ALL) konkordanssituloksiin, koska osumia sillä oli lähes 5 930. Selvä epätyypillisuus on puolestaan esimerkin (57) ilmaus *mennä hermoille*, sillä konkordanssihaku yhdistelmällä MENNÄ & HERMO (ALL) ei antanut yhtään osumaa. Sen sijaan kollokaatio-osumia oli hakuyhdistelmällä KÄYDÄ & HERMO (ALL) yli 80, eli *käydä hermoille* on selvästi tyypillisempi ilmaus.

- (56) *Ihmiset eivät haluaa **lähteä** (pro **jäädä**) eläkkeelle.* (VE0159c)
- (57) *Netti **menee** (pro **käy**) hermoille* (UN0004)

Myös adverbien yhteydessä on tehty epätyypillisiä verbinvalintoja. Ilmaus *ottaa vakavasti* on natiivisuomessa vakiintunut, mikä näkyy hakuyhdistelmän OTTAA & VAKAVASTI konkordanssitulosten 3 230 osumasta. *Ottaa*-verbin sijaan on käytetty verbiä *kohdella* esimerkissä (58) ja verbiä *käsitellä* esimerkissä (59). Kumpikaan tapaus ei liene tyypillinen, sillä kollokaatio-osumia oli hakuyhdistelmällä KOHDELLA & VAKAVASTI vain yksi ja hakuyhdistelmällä KÄSITELLÄ & VAKAVASTI kymmenkunta.

- (58) *Se on lempiharrastuksemme ja sitä ei voi **kohdella** (pro **ottaa**) ihan vakavasti.* (PU0002g)
- (59) *Siksi on minulle niin tärkeä, että koko Puolassa **käsitellään** (pro **otetaan**) ympäristönsuojelua vakavasti.* (PU0032)

4.2.4. Semanttinen preferenssi ja semanttinen prosodia

Semanttisen preferenssin ja semanttisen prosodian alaryhmän tapaukset eroavat yksittäisistä lähimerkityksistä sanoista ja kollokaatioista siinä, että epätyypilliseen myötäesiintymään liittyy myös semanttinen ulottuvuus (ks. luku 3.2.3.3. ja 3.2.3.4.). Sanaparin vaihtoehtojen semanttisen preferenssin tai semanttisen prosodian saa selville analysoimalla niiden konkordanssituloksia kvalitatiivisesti. Tulokset antavat kuvan siitä, millaisten semanttisten ryhmien kanssa vaihtoehdot tyypillisesti esiintyvät ja millainen merkityssävy (positiivinen, negatiivinen vai neutraali) sanan sisältävissä virkkeissä yleensä valitsee. Alaryhmä jakautuu kolmeen keskittymään, jotka ovat epätyypillinen semanttinen preferenssi adjektiiveissa, epätyypillinen semanttinen preferenssi verbeissä sekä epätyypillinen semanttinen prosodia.

Epätyypillinen semanttinen preferenssi adjektiiveissa. Esimerkissä (60) adjektiivia *vaivalloinen* on käytetty määrittämään elollista tarkoitetta. Vaikka Kielitoimiston sanakirja antaa yhdeksi käyttöesimerkiksi *vaivalloinen potilas, asiakas* (KS s.v. *vaivalloinen*), tällainen käyttö lieenee käytännössä suhteellisen harvinaista, sillä hakusanan VAIVALLOINEN yli 290 osumaa kertoo sanan selvästä semanttisesta preferenssistä elottomiin tarkoitteisiin. Sen sijaan adjektiivi *hankala* näyttää olevan yleisessä käytössä elollisen tarkoitteen määritteenä, sillä useita osumia löytyy esimerkiksi hakuyhdistelmillä HANKALA + IHMINEN (yli 50), HANKALA + ASIAKAS (yli 40) ja HANKALA + TYÖNTEKIJÄ (yli 20).

- (60) *Muumipapalle oli ollut kerrottu, että tähtien mukaan (hemulit uskoivat nimenomaan astrologiaan) hänestä tulee **vaivalloinen** (pro **hankala**) muumi ylilajakkuudestaan vuoksi.* (PU0004k)

Myös esimerkissä (61) adjektiivia *valkea* on käytetty ihmisten kuvaamiseen, vaikka se yleensä määrittää elotonta tarkoitetta tai elollista oliota, joka ei ole ihminen. Hakuyhdistelmän VALKEA + SUBST yli 1 060 osumasta ilmenee, että *valkea*-adjektiivilla on

semanttinen preferenssi erityisesti elottomiin tarkoitteisiin (esim. VALKEA + JOULU, AINE, TALO) sekä pienemmissä määrin eläimiin (esim. VALKEA + LINTU, RATSU) ja kasveihin (esim. VALKEA + RUUSU). Sen sijaan hakuyhdistelmän KALPEA + SUBST yli 140 osumasta selviää, että adjektiivilla on selvä semanttinen preferenssi ihmisiin viittaaviin tarkoitteisiin (esim. KALPEA + NAAMA, SOTILAS, POIKA).

- (61) *Kuolemankauhusta **valkeat** (pro **kalpeat**) ihmiset poistuivat bussista.* (PU0005p)

Epätyypillinen semanttinen preferenssi verbeissä. Esimerkeissä (62) ja (63) verbiä *lahjoittaa* on käytetty lauseessa, jossa objekti on *vapaus* tai *vapaapäiviä*. Hakuyhdistelmän LAHJOITTAA & OBJ yli 5 870 osumasta kuitenkin ilmenee, että verbin semanttinen preferenssi on vahvimmillaan rahaan tai konkreettisiin esineisiin viittaaviin tarkoitteisiin, ja sitä käytetään huomattavasti harvemmin abstraktien tarkoitteiden yhteydessä. Lisäksi hakuyhdistelmä LAHJOITTAA & VAPA* (OBJ) ei antanut yhtään osumaa, jossa lahjoitettavana olisi *vapaus* tai *vapaapäivä*. Yksi osuma tosin oli, jossa lahjoitettiin *vapaa-aikaa*.

- (62) *Yhtäaikaa Katariinalle sanottiin kuninkaan Eerikin voivan **lahjoittaa** (pro **antaa** ~ **myöntää** ~ **suoda**) hänelle vapautta.* (PU0001e)
- (63) *Toivon, että yliopisto **lahjoittaa** (pro **antaa**) meille vapaa päiviä marraskuun alussa.* (PU0022)

Esimerkissä (64) verbiä *pilaantua* on käytetty lauseessa, jonka subjekti on abstrakti tarkoite. Hakusanan PILAANTUA 2 280 osumasta kuitenkin selviää, että verbiä käytetään ylivoimaisesti konkreettisten, 'fyysistä ainesta' merkitsevien sanojen yhteydessä. Lisäksi konkordanssihakuyhdistelmällä PILAANTUA & TILANNE (SUBJ) ei antanut yhtään osumaa. Sopivampi verbi tässä tilanteessa olisi siis *huonontua* tai *huonota*.

- (64) *Kun tilanne **pilaantui** (pro **huonontui** ~ **huononi**), se tapahtui suuressa määrin Markun syystä.* (PU0005k)

Esimerkissä (65) verbiä *karkottaa* on käytetty elottoman ja abstraktin objektin yhteydessä ja esimerkissä (66) verbiä *osallistua* on käytetty elottoman subjektin yhteydessä. Kumpikaan tapaus ei liene tyypillinen, sillä hakusanan KARKOTTAA lähes 4 270 osuman perusteella karkotettava kohde on useimmiten ihminen tai eläin. Samoin hakusanan

OSALLISTUA yli 77 220 osumaa antaa ymmärtää, että osallistuja on useimmiten ihminen tai ihmisistä koostuva entiteetti. Tyypillisempi vaihtoehto olisi *hävittää* esimerkissä (65) ja *sisältyä* esimerkissä (66), sillä hakusanan HÄVITTÄÄ lähes 5 650 osumassa ja hakusanan SISÄLTÄÄ yli 5 210 osumassa esiintyy runsaasti elotonta tarkoitetta merkitseviä tapauksia.

- (65) *Se osoittaa, että Neuvostoliiton valta yrittää **karkottaa** (pro **hävittää**) suomalaisuutta jopa ruokalistalta.* (PU0006e)
- (66) *Infinitiivikollokaatioiden vertailuun **osallistuivat** (pro **sisältyivät**) vain ne kollokaatiot, jotka löysin ryhmittelyn jälkeen aineistostani.* (VI0393b)

Epätyypillinen semanttinen prosodia. Epätyypillisen semanttisen prosodian tapauksia ei ole paljon aineistossani. Niistä yleisimmät ovat substantiivin *elämys* käyttö negatiivisessa yhteydessä ja verbin *joutua* käyttö positiivisessa yhteydessä. Kielitoimiston sanakirjan mukaan *elämys* tarkoittaa '(voimakkaasti vaikuttavaa) kokemusta', kun taas *kokemus*-sanalla on neutraalimpi merkitys 'koettu tapaus' (KS s.v. *elämys*, *kokemus*). Esimerkeissä (67) ja (68) sanaa *elämys* on käytetty viittaamaan negatiiviseen kokemukseen, mikä on ilmeistä sanojen *huonosti* ja *kauhea* käytöstä. Hakusanan ELÄMYS lähes 3 300 osumasta kuitenkin selviää, että sana esiintyy ylivoimaisesti positiivisissa yhteyksissä eli sillä on positiivinen semanttinen prosodia.

- (67) *Seuraava Suomisen **elämys** (pro **kokemus**), joka vahvitti käsityksen, kuinka huonosti toimi keskeisesti suunniteltu kommunistinen talous ja kuinka jahjoittava sen yhteiskunta oli, on tapahtuma, kun discon jälkeen Suominen haluaa ottaa taksin.* (PU0006e)
- (68) *Ensin ikävystyneenä, mutta sitten enemmän ja enemmän kinnostuneena, novellin kertoja kuuntelee naisen muistelevan erilaisia vanhoja, usein kauheita **elämyksiään** (pro **kokemuksiaan**).* (PU0032f)

Esimerkeissä (69–71) verbiä *joutua* on käytetty positiivisessa tai neutraalissa yhteydessä. Kielitoimiston sanakirjan mukaan sanaa käytetään kuitenkin muun muassa ilmaisemaan ”jtk kohtaavaa tahatonta (epäedullista) muutosta” (KS s.v. *joutua*), eli sillä on tyypillisesti negatiivinen semanttinen prosodia. Myös hakusanan JOUTUA 160 310 osumasta selviää, että sana esiintyy lähinnä negatiivisissa yhteyksissä.

- (69) *Sitten me **joudumme** (pro **päädymme**) [televisiosarjan jakson] ensimmäiseen kohtaukseen.* (HO0001k)
- (70) *Hänen täytyy puolustautua oikeudessa, kun hän on **joutunut** (pro **päässyt**) taas sairaalasta.* (SA0107i)
- (71) *Vaikka se kuvattiin ennen Paha Maata niin valkokankaalle **joutuikin** (pro **pääsi**) se myöhemmin.* (VI0346a)

4.2.5. Uudismuodosteet

Kuuluvaiselta (2015) lainaamani **uudismuodosteiden** alaryhmä koostuu tapauksista, joissa suomenoppija on tavoitellut jotakin tiettyä sanaa tai ilmausta mutta päätynyt kuitenkin tulokseen, joka on semanttisesti ymmärrettävä mutta jollakin tavalla natiivisuomesta poikkeava. Kuuluvaisen (mts. 38) tapaan lasken uudismuodosteiksi vain ne tapaukset, jotka noudattavat kirjoitetun yleiskielen morfologiaa. Alaryhmä jakautuu kahden keskittymään, yksi- ja useampisanaisiin uudismuodosteisiin.

Yksisanaiset uudismuodosteet. Osa yksisanaisista uudismuodosteista koostuu tapauksista, joissa sekä epätyypillinen muoto että sen tyypillisempi vaihtoehto ovat yksisanaiset. Tapausten muodostamisperusteet ovat useimmiten selkeästi jäljitettävissä. Esimerkissä (72) suomenoppija on luultavasti muodostanut refleksiivisen muuttumisjohdoksen (ks. VISK § 335) *paimentaa*-verbistä tietämättä, että ilmiölle on jo oma verbi *laiduntaa*. Esimerkin (73) *kirjoituskieli* on puolestaan jo hyvin lähellä natiivisuomen muotoa. Esimerkissä (74) taas näkyy suomenoppijan äidinkielen vaikutus sanan muodostamiseen, sillä yksi viron *turg*-sanon merkityksistä onkin 'markkinat' (EKI ühendsõnastik s.v. *turg*).

- (72) *Pietrek näki kuin kaksi vihaista lihavaa pukkia **paimentui** (pro **laidunsi**) niityllä.* (PU0025b)
- (73) *En keskustellut paljon suomalaisten kanssa suomeksi, koska olin aloittelija, ja tunneilla opettaja puhui suomen kieliopista tavallisesti ja hän yritti puhua suomen **kirjoituskieltä** (pro **kirjakieltä**), koska olimme ajoittelijat ja vain siis ymmärsimme häntä.* (UN0013b)
- (74) *Nykytilanne Viron **työturulla** (pro **työmarkkinoilla**) on se, että jos ei osaa venäjää niin töitä helposti ei löydy.* (VI0399a)

Loput yksisanaista uudismuodosteista muodostavat tapaukset, joissa yhdensanaisen uudismuodosteen tyypillisempi vaihtoehto koostuu useammasta sanasta. Myös näiden tapauksien muodostamisperusteet ovat useimmiten helposti arvattavissa. Esimerkin (75) kaltaisia *finanssi*-alkuisia yhdyssanoja löytyy jossain määrin Kielitoimiston sanakirjasta (KS s.v. *finanssi*-), joskaan *finanssiongelma* ei kuulu niiden joukkoon. Tyypillisyyden ja epätyypillisyyden rajalla on taas esimerkin (76) *itsemurhaaja*, joka on sen verran vakiintunut, että sille löytyy oma sana-artikkeli Kielitoimiston sanakirjasta (KS s.v. *itsemurhaaja*). Vaikka hakuyhdistelmällä ITSEMURHA + TEKIJÄ on kaksinkertaisesti osumia (12) verrattuna hakusanaan ITSEMURHAAJA (6), summat ovat niin pienet, että mitään varmoja johtopäätöksiä ei kannata niiden perusteella tehdä. Sana ei kuitenkaan sopine esimerkin (76) kontekstiin, sillä *itsemurhaaja* viittaa henkilöön, joka on jo suorittanut teon.

(75) *Pian tulee finanssiongelmiakin* (pro *taloudellisiakin ongelmia*). (PU0007d)

(76) *Hämmästyksen ladossa Rellonen tapaa toisen itsemurhajan* (pro *itsemurhan yrittäjän*), eversti Hermannin Kemppaisen. (PU0005o)

Useampisanaiset uudismuodosteet. Raja useampisanaisten uudismuodosteiden ja lauseketason epätyypillisyyksien (ks. luku 4.4.) välillä voi joskus olla häilyvä. Suurin osa tämän keskittymän tapauksista on sellaisia, jotka voidaan korvata natiivisuomessa yhdellä sanalla. Yksisanaisten uudismuodosteiden tavoin myös useampisanaisten tapauksien muodostamisen takana on usein selkeä logiikka. Esimerkkien (77) ja (78) tapaukset ovat jo hyvin lähellä natiivisuomen muotoa, kun taas esimerkissä (79) on käytetty *ongelma*-sanan harvinaisempaa synonyymia. Esimerkin (80) taustalla on luultavasti suomenoppijan äidinkielen eli viron vaikutus (ks. EKI ühendsõnastik s.v. *vastu*), ja esimerkissä (81) on puolestaan keksitty uudenlainen nimi luonnonilmiölle.

(77) *Ensimmäisen vaikutuksen* (pro *ensivaikutelman*) jälkeen tuli kuitenkin ajatteleminen ja miettimään pakottaminen on epäilemättä sen elokuvan hyvä puoli. (PO0005f)

(78) *Kieli on jotakin, joka tekee ihmisestä yhteiskunnallisesti kelpaavan* (pro *yhteiskuntakelpoisen*). (TS0014e)

(79) *Luultavasti se tapahtuu siitä syystä, että kielellisissä pulmallisissa tilanteissa* (pro *ongelmatilanteissa*) ihmisen aivot automaattisesti poimivat tietoja eniten käytetystä tilasta eli siitä paikasta, jossa on tietoa äidinkielestä ja sen piirteistä. (VE0172)

- (80) *Häntä järkyttää näky niin kovasti, että hän ottaa töistä vapaata ja aloittaa **ma-sennuksen vastaisten pillerien** (pro **masennuspillerien**) käyttämisen.* (VI0346)
- (81) *Siksikin runoista, lauluista ja muista teksteistä löytyy niin paljon kuvauksia **valkoisista öistä** (pro **yöttömistä öistä**) ja ilta-auringosta.* (PU0022g)

4.2.6. Sanakirjan sudenkuopat ja transfer-tapaukset

Sanakirjan sudenkuoppien ja transfer-tapausten alaryhmä koostuu tapauksista, joiden muodostamisperusteet ovat jollain tavoin selitettävissä siitä huolimatta, että ne eivät ole suomeksi ajateltuna semanttisesti järkeenkäypiä. Uudismuodosteiden tapaan nämäkin epätyypillisyydet alkavat jo lähestyä virheitä, sillä ne saattavat olla semanttisesti harhaanjohtavia eikä niitä välttämättä hyväksytä nyky-yleiskielen normien puitteissa. **Sanakirjan sudenkuopiksi** kutsun niitä tapauksia, jotka syntyvät luultavasti sanakirjan ”väärinkäytöstä”. Suomenoppija on siis hakenut kaksikielisestä sanakirjasta tiettyä sanaa toisella kielellä ja valinnut suomenkielisten tulosten listalta käyttöyhteyteen sopimattoman vaihtoehdon, joka on jo merkitykseltään kaukana tyypillisemmästä vaihtoehdosta. **Transfer-tapaukset** ovat puolestaan epätyypillisiä sananvalintoja, joissa suomenoppijan äidinkielen tai muun osaaman kielen vaikutus on selvästi havaittavissa.

Sanojen liian ja kovin sekoittaminen. Sanakirjan sudenkuopat jakautuvat viiteen keskittymään, joista ensimmäinen on *liian*-sanan käyttö *kovin*-sanan tilalla merkityksessä ’(ei) järin’. Esimerkeissä (82) ja (83) käytetty EI + ... + LIIAN + ADJ/ADV -rakenne muistuttaa muun muassa englannin NOT + TOO + ADJ/ADV -rakennetta, jonka merkitys on ’ei kovin ...’. Yksi *too*-sanan suomenkielisistä käännöksistä onkin *liian* (Sanakirja.org englanti–suomi s.v. *too*).

- (82) *Puolalaiset eivät ehkä yleisesti **liian** (pro **kovin**) älykkäitä, mutta eivät he niin typeriä olekaan!* (PU0001d)
- (83) *Vaikka suomen kieli on kolmantena kielenä vironkielisissä kouluissa ei **liian** (pro **kovin**) usein valittu kolmas kieli, on kuitenkin opiskelijoita yhteensä noin tuhansi ja verrattuna viron väkilukuun se ei ole **liian** (pro **kovin**) huono saavutus.* (VI0392c)

Fokuspartikkelien *jopa* ja *edes* sekoittaminen. Toinen keskittymä koostuu tapauksista, joissa fokuspartikkelit (ks. VISK § 839) *jopa* ja *edes* on sekoitettu keskenään. Vaikka kumpikin sana voidaan kääntää englanniksi sanalla *even* (Sanakirja.org englanti–suomi s.v. *even*), ne muodostavat polaarisen parin (ks. VISK § 1635), jossa *jopa* esiintyy myönteisissä yhteyksissä ja *edes* esiintyy kieltolauseissa ja muissa kielteisissä yhteyksissä (VISK § 849). Esimerkeissä (84) ja (85) *jopa*-sanaa on käytetty kielteisessä lauseessa, vaikka tyypillisempi vaihtoehto olisi *edes*. Esimerkissä (86) on puolestaan käytetty *edes*-sanaa myönteisessä lauseessa, vaikka tyypillisempi vaihtoehto olisi *jopa*.

- (84) *Jotenkin, bussissa oli niin paljon ihmisiä, että ei voitaisi tunkea sisälle jopa (pro edes) neulaa.* (PU0001d)
- (85) *Jopa (pro edes) kun Markulla on lopulta rahaa, hän ei osaa käyttää sitä hyödyllisellä tavalla.* (PU0005k)
- (86) *Ensimmäiset pari-kolme viikkoa tai edes (pro jopa) koko kuukausi, kun elin Suomessa yritin opetella sanoja, ymmärtää kielen rakennetta, taivutusjärjelmiä, jotta ainakin ymmäraisin suomalaisten puhetta.* (VE0172)

Sanojen *siis* ja *joten* sekoittaminen. Kolmas keskittymä koostuu tapauksista, joissa partikkelia *siis* on käytetty relatiivisen proadverbin *joten* tilalla. Vaikka molemmat sanat voidaan kääntää englanniksi sanalla *so* (Sanakirja.org englanti–suomi s.v. *so*), ne eivät aina ole vaihdettavissa keskenään suomessa. Esimerkeissä (87) ja (88) partikkelia *siis* on käytetty kausaalisen eli syy–seuraussuhteen merkitsimenä. Tällainen käyttö on kuitenkin epätyypillistä, sillä *siis* ei yleensä esiinny lauseen alussa silloin, kun sitä käytetään kahden lauseen konnektiivina päätelmien ja seurausten ilmaisemiseksi (ks. VISK § 1128, § 1132). Myös hakusanan SIIS (VIRKKEEN ALUSSA) yli 2 270 osumasta ilmenee, että sana esiintyy lauseen alussa lähinnä silloin, kun sitä käytetään esimerkiksi suunnittelupartikkelina itsekorjauksissa (ks. VISK § 861, § 862). Sen sijaan relatiivinen proadverbi *joten* esiintyy aina lauseen alussa (VISK § 1132) ja sopisi paremmin seuraavien esimerkkien kontekstiin.

- (87) *Kasia oli ollut velkaa mulle jotakin päivällistä, siis (pro joten) tänään päätin käyttää sitä hyväkseni ja keitin pinaattimakaronia (tietysti pinaatti oli Kasian oma).* (PU0031)
- (88) *Työkulttuuri on koko kansan kulttuurin osa, siis (pro joten) vieraan maan työkulttuuriakin on osattava.* (VE0095h)

Yleis- ja erikoiskielen sanojen sekoittaminen. Neljäs keskittymä koostuu tapauksista, joissa sanan suomenkielisissä käännöksissä on vaihtoehtoina sekä jokin yleisessä käytössä oleva sana että spesifisempi, yleensä erikoiskieleen kuuluva termi. Epätyypillisuus syntyy useimmiten siitä, että erikoiskielen termiä on käytetty sille sopimattomassa kontekstissa, kuten esimerkeissä (89–93). Kaikissa tapauksissa sanojen konkordanssitulokset (hakusanalla RELATIIVINEN vs. SUHTEELLINEN, RESONANSSI vs. VASTAKAIKU jne.) antavat ymmärtää, että yleiskielen sana on ylivoimaisesti suositumpi muissa kuin erikoisalan yhteyksissä.

- (89) *Mikäli Suomesta ja suomalaisista kirjoittaminen oli **relatiivisen** (pro **suhteellisen**) helppo, sikäli Puolaa ja puolalaisia kuvaaminen tekee vaikeuksia.* (PU0003h)
- (90) *Sekä paperista että huopakankaasta tehtyjä joulullisia tounteita voi myös ostaa seuran joulumarkkinoilta, jotka saavat joka vuosi kävijöiltä positiivista **resonanssia** (pro **vastakaikua**).* (SA0098c)
- (91) *Kotona sitten, jossa olen rauhassa omin päin, voin käyttää aikaa **kreatiiviseen** (pro **luovaan**) toimintaan kuin esim. edellä mainitun kääntämiseen.* (TS0006j)
- (92) *Kieli on myös merkittävä **faktori** (pro **tekijä**).* (VE0093i)
- (93) *Siitä syystä suomen kielen toisena ja vieraana kielenä opettamisen ja oppimisen tutkimus on hyvin **aktuaalinen** (pro **ajankohtaista**).* (VE0172a)

Myös sanojen sekoittamista toiseen suuntaan tapahtuu, joskin huomattavasti harvemmin. Esimerkissä (94) on käytetty ilmausta *kolmas ihminen* viittaamaan kieliopilliseen kategoriaan, jolloin tyypillisempi ilmaus on *kolmas persoona* (KS s.v. *persoona*).

- (94) *Minna Canth kirjoittaa Salakarin kolmannen **ihmisen** (pro **persoonan**) näkökulmasta.* (KI0012)

Yksittäiset sopimattomat sanat. Viides keskittymä koostuu tapauksista, joissa sanaparien molemmat osat ovat jo semanttisesti melko kaukana toisistaan, vaikka ne olisivatkin käännettävissä samaksi sanaksi jollakin toisella kielellä. Esimerkkien (95–99) epätyypillisyydet ovat selitettävissä Sanakirja.org -kaltaisen verkkosanakirjan avulla. Sen mukaan hakusanan *przytulny* suomenkielisissä tuloksissa esiintyy sekä *kodikas* että *mukava*, hakusanan *przyznać* tuloksissa esiintyy sekä *myöntää* että *tunnustaa*, hakusanan *los* tuloksissa esiintyy sekä *arpa* että *kohtalo*, hakusanan *orange* tuloksissa esiintyy sekä *appelsiini*

että *oranssi*, ja hakusanan *miss* tuloksissa esiintyy sekä *puuttua* että *kaivata* (Sanakirja.org puola–suomi s.v. *przytulny, przyznać, los*; Sanakirja.org englanti–suomi s.v. *orange, miss*).

- (95) *Ja ne, joiden täytyi seisoa, ei voineet jopa seisoa **kodikkaasti** (pro **mukavasti**) eli suorana, koska autokuljettaja ei ollut viitsinyt avata luukuja ja sitten kaikkien matkatavaroiden täytyttiin ottaa sisälle ja ihmiset seisoivat niiden päällä, alla, vieressä ja ympärillä.* (PU0001d)
- (96) *Seura on hakee rahoitusta projektilleen Lääninhallituksesta, joka **tunnustaa** (pro **myöntää**) rahaa Euroopan aluekehitysrahastosta (EAKR).* (PU0003a)
- (97) *Sen esimerkki on Puolan lippu ja kansallisvärit, jotka siinä näkyy eli **arpaa** (pro **kohtaloa**) merkitsevä valkoinen ja punainen, joka on veren symboli.* (PU0027k)
- (98) *Muistan, että minulle lapsuudestani saakka aivan käsitlemätöntä oli, että kommunismin aikana kauppojen hyllyt olivat tyhjiä tai että **oranssit** (pro **ap-pelsiinit**) ja sitruunat olivat saatavissa vain ennen joulua ja vielä vain silloin, kun seisoj jonossa riittävästi kauan (mikä tarkoitti usein koko yön).* (PU0004b)
- (99) *Joskus **minulta puuttui** (pro **kaipasin**) asioita, joita olimme tehneet yhdessä.* (PU0015c)

Transfer-tapaukset. Seuraavat esimerkit sisältävät epätyypilliset sananvalinnat tai rakenteet, jotka ovat selvästi peräisin suomenoppijan äidinkielestä. Esimerkissä (100) *ehkä* on käännetty suoraan venäjän vastineesta *может быть* (MOT Venäjä s.v. *может быть*). Esimerkeissä (101–104) taas näkyy viron vaikutus, sillä ’vaimo’ on viroksi *naine*, ’jotta’ ilmaistaan konjunktiolla *et*, ’huono puoli’ on *miinus* ja ’sokea’ on *pime* (EKI ühendsõnastik s.v. *naine, miinus, pime*; MOT Kotus Suomi–viro-suursanakirja s.v. *et*).

- (100) *He ovat sotaretkellä jossain idässä (**voi olla** (pro **ehkä**) Suomen rajaseudulla tai jopa Karjalassa), siis varmasti tapahtumat ajoittuvat Talvisodan aikaan (sotatematiikkaan viittaa Merin käyttämä sanasto: sotaretki, korsu, sotamies, tykkiasemat).* (VE0129)
- (101) *Hän oli omistautunut perheelle, ei ollut väkivaltainen, ei juopotellut ja sieti senkin, että **nainen** (pro **vaimo**) oli lähtenyt kahdeksi kuukaudeksi rakastajattarensa luo Ranskaan.* (VI0346a)
- (102) *Opettajan on autettava oppilas kielikynnyksen yli, **että** (pro **jotta**) hän pärjäisi jatkossa omin voimin - kielihan on elämänavalue, jolla oppiminen ei lopu koskaan.* (VI0395)

(103) *Minun **miinukseni** (pro **huono puoleni**) on se että en heti ymmärrä mistä puhutaan, mutta ihmisiä on erilaisia: jotkut osaavat kuunnella ja jotkut osaavat lukea, kolmannet taas puhua.* (VI0399a)

(104) *Tai olenko itse niin **pimeä** (pro **sokea**), että en näe tosiasiaa?* (VI0399a)

4.2.7. Virheelliset nominivalinnat

Virheellisten nominivalintojen alaryhmään kuuluvat tapaukset, joissa suomenoppijan valitsema nominijohdos ei sovi semanttisesti virkkeen kontekstiin. Ryhmän sanaparien vaihtoehdot johdetaan useimmiten samasta kantasanasta, ja johdosten merkityksetkin ovat monesti hyvin samanlaisia. Osaa tapauksista voitaneen kuitenkin jo pitää virheinä, sillä suomenoppijan valitsema johdos välittää usein eri merkityksen kuin hänen alun perin oletettavasti tarkoittamansa.

Epätyypilliset tai virheelliset substantiivinvalinnat. Esimerkeissä (105) ja (106) sanaparien vaihtoehtojen välillä on hienoinen merkitysero. Kielitoimiston sanakirjan mukaan *vaihtelu* tarkoittaa muun muassa 'vaihtumista', vaikka virkkeen asiayhteydessä kyse on pikemminkin 'vaihtamisesta' (KS s.v. *vaihtelu*, *vaihto*). Samalla *ajattelu* viittaa 'ajattelemiseen' tai 'ajattelutapaan' eikä 'ajattelun tuotteeseen', joka sopisi virkkeeseen paremmin (KS s.v. *ajattelu*, *ajatus*). Esimerkissä (107) vaihtoehtojen merkitykset ovat taas sen verran kaukana toisistaan, että suomenoppijan valitsema *valmistus* ei enää sovi lainkaan virkkeen kontekstiin, sillä siltä puuttuu 'tutkimuksen suorittamisen' merkitys (KS s.v. *valmistus*, *valmistua*).

(105) *Lyhyen mielipiteiden **vaihtelun** (pro **vaihdon**) jälkeen he tekevät kompromissin, jonka mukaan osa julisteiden iskulauseita tulee olemaan väreissä ja toinen mustavalkoista.* (PU0003k)

(106) *Toinen **ajattelu** (pro **ajatus**) joka vain vähän liittyy tähän artikkeliin oli miksi Suomussalmi.* (PU0018)

(107) *Sen lisäksi, Suomi on Viron lähinaapuri, joten **valmistuksen** (pro **valmistumisen**) jälkeen, kun minä jo osaan suomea, voisin saada vaikkapa Helsingistä hyvän työpaikan.* (VE0173a)

Epätyypilliset tai virheelliset adjektiivivalinnat. Esimerkkien (108) ja (109) tapauksissa kyse on lähinnä epätyypillisyydestä, sillä sanaparien vaihtoehdot viittaavat samaan tarkoitteeseen. Pääero vaihtoehtojen välillä lienee niiden käyttötaajuus, sillä *kade*-sanankäyttöä merkityksessä 'kateellinen' pidetään harvinaisena, ja *perinteellinen*-sanaa tavallisempaan pidetään sen synonyymia *perinteinen* (KS s.v. *kade*, *perinteellinen*). Merkitysero esimerkin (110) vaihtoehtojen välillä on puolestaan suurempi, sillä *ammattillinen* tarkoittaa 'ammattia tai ammatteja koskevaa', vaikka virkkeen kontekstiin sopisi paremmin merkitys 'ammatinharjoittajan tapaan toimiva' (KS s.v. *ammattillinen*, *ammattimainen*).

(108) *Todettiin, että Suomessa on raskasta elää, suomalaiset ovat toisilleen julmia ja kateita* (pro *kateellisia*), *voimakkaat käyttävät hyväkseen heikkoja, luontoa myrkytetään, työelämä tuntuu oravanpyörältä eikä ole olemassa hyvää tapaa parantaa oloaan.* (PU0005p)

(109) *Wrangelsburgilainen Papier und Druckin seura on vuonna 2002 perustettu järjestö, joka tarjoaa kiinnostuneille mahdollisuuden tuottaa luonteellisista aineistosta paperia perinteellisellä* (pro *perinteisellä*) *tavalla.* (SA0098c)

(110) *Konsertin arvostelu on hyvin ammatillinen* (pro *ammattimainen*), *artikkelistä löytyy paljon aiheeseen liittyviä yksityiskohtia.* (PU0022e)

4.2.8. Virheelliset verbinvalinnat

Myös **virheellisten verbivalintojen** alaryhmä koostuu tapauksista, joiden sanaparit muodostavat samasta kantasanasta johdetut vaihtoehdot. Erona edelliseen ryhmään on kuitenkin se, että useimmissa tapauksissa suomenoppijan valitsema verbijohdos on suorastaan virheellinen ja lopputulos ei ole enää kieliopillisesti tai semanttisesti hyväksyttävä. Vaikka keskitynkin tässä tutkimuksessa nimenomaan epätyypillisyyksiin, virheellisiä verbinvalintoja esiintyy aineistossani sen verran paljon, että pidän ilmiötä mainitsemisen arvoisena.

Transitiivi- ja intransitiiviverbien sekoittaminen. Ryhmän suurin keskittymä muodostuu tapauksista, joissa transitiivi- ja intransitiiviverbit (ks. esim. VISK § 459, § 461) sekoitetaan keskenään. Yksi syy tähän sekoittumiseen voi olla se, että sanakirjahaun suomenkielisissä tuloksissa on usein lueteltu molemmat verbit ilman sen tarkempaa tietoa niiden käyttöyhteyksistä. Aineistoni yleisin tapaus on verbien *aloittaa* ja *alkaa*

sekoittaminen. Esimerkissä (111) intransitiivista *alkaa*-verbiä on käytetty transitiivisen *aloittaa*-verbin tavoin:

- (111) *Tasly **alkoi** (pro **aloitti**) yleisen kansainvälistymisen huhtikuussa 2006.* (KI0009h)

Tällaisia tapauksia, joissa *alkaa*-verbiä on käytetty *aloittaa*-verbin tilalla, on vain yksi. Sen sijaan paljon yleisempiä ovat tapaukset, joissa *aloittaa*-verbiä on käytetty *alkaa*-verbin tilalla kuten esimerkeissä (112) ja (113):

- (112) *Kuitenkin, torstaisin tuntini **aloittavat** (pro **alkoivat**) suhteellisesti myöhän, siis kun me keskusteltiin eilen puhelimesta pyysin äitiäni, että hän ei soitti tänään koska haluaisin nukkua vähän kauemmin.* (PU0001d)

- (113) *Ja nyt kun syksy **aloitti** (pro **alkoi**) ja ilma muutti rumaksi ja kylmäksi, tarvitsee aikaa kun voi rentoutua ja varastoida voimaa talveksi.* (PU0026)

Esimerkeissä (114) ja (115) on puolestaan käytetty ALOITTA + A-INF -rakennetta virheellisesti infinitiivilausekkeen tapaan (ks. VISK § 497), jolloin A-infinitiivi toimii *aloittaa*-verbin objektina. Tämä ei kuitenkaan ole mahdollista, sillä *aloittaa*-verbin vaatima objekti ei voi olla A-infinitiivimuotoinen verbi. Sen sijaan oikea rakenne olisi verbiketju (ks. VISK § 496) ALKAA + A-INF.

- (114) *Jacek **aloitti** (pro **alkoi**) elää rikkaana hallitsijana, jonka jokainen oikku täyttettiin.* (PU0005e)

- (115) *Olen **aloittanut** (pro **alkanut**) lukemaan M. Waltarin ”Sinuhe Egyptiläistä” vielä viime vuonna.* (VE0100)

Toiseksi yleisin tapaus on verbien *keskittää* ja *keskittyä* sekoittaminen. Vaikka kumpikin verbi voi esiintyä illatiivimuotoisen substantiivilausekkeen yhteydessä, niiden oikea valenssi on jäänyt huomioimatta seuraavissa esimerkeissä. Verbi *keskittyä* on kaksipaikkainen (ks. VISK § 446), eli sillä on kaksi täydennystä, subjekti ja illatiivimuotoinen substantiivi- tai MA-infinitiivilauseke. Sen sijaan *keskittää*-verbin täydennyksiin kuuluvat subjekti, objekti ja useimmiten myös kolmas substantiivilauseke. Esimerkeissä (116) ja (117) *keskittää*-verbin objekti puuttuu, kun taas esimerkin (118) verbinvalinta ja puuttuva substantiivilauseke tekevät virkkeestä semanttisesti epä johdonmukaisen.

- (116) *Yritän tehdä kaikki ajoissa, mutta tiedän jo etten ehdi, kun en osaa **keskittää** (pro **keskittyä**) opiskeluun...* (PU0011d)
- (117) *Olen kuitenkin unohtanut, että ilman joitakin perushyvyksiä ei pysty **keskittämään** (pro **keskittymään**) sisäiseen elämään ja sen tarpeisiin.* (PU0027i)
- (118) *Kertoessaan omasta historiasta he aina **keskittävät epäreiluja kohteluja** (pro **keskittävät epäreiluihin kohteluihin**), jotka heille tehtiin.* (PU0027k)

Myös transitiiviverbi *kehittää* ja intransitiiviverbi *kehittyä* on sekoitettu keskenään. Kaikissa tapauksessa transitiivista *kehittää*-verbiä on käytetty intransitiivisen *kehittyä*-verbin tilalla, kuten esimerkeissä (119) ja (120):

- (119) *Huhutaan, että saksalaisen kebabin kuuluisusta maailman ympärillä **kehitti** (pro **kehittyi**) niin, koska kaupunkilaiset haluavat syödä nopeasti ja kävellessä ja siksi se on heistä parempaa ratkaisua.* (SA0106e)
- (120) *Heidän mielikuvansa on **kehitetty** (pro **kehittynyt**) hyvin.* (VE0130a)

Yllä mainittujen tapausten lisäksi aineistossani on myös yksittäisiä tapauksia, joissa transitiivi- ja intransitiiviverbit on sekoitettu keskenään. Yleisempiä ovat esimerkkien (121) ja (122) kaltaiset tapaukset, joissa transitiiviverbiä on käytetty lauseessa, jossa olisi pitänyt käyttää intransitiiviverbiä. Myös esimerkkien (123) ja (124) kaltaisia tapauksia, joissa intransitiiviverbiä on virheellisesti käytetty transitiiviverbiä edellyttävissä lauseissa, esiintyy jonkin verran.

- (121) *Vähintään niin kauan kuin asuntomme ei **polta** (pro **pala**) poroksi, mutta sellaisessa tapauksessa analogisetkin muistomme katoavat...* (PU0001c)
- (122) *Runon sisältö ei **avaa** (pro **avaudu**) (eli sen merkitys ei ilmene selvästi), vaan runon muoto on etualalla.* (SA0032)
- (123) *Ensin hän irvailee, että tilanne osoittaa Kiviniemen osaavan **kietoutua** (pro **kietoa**) muut sormen ympärille.* (PU0007)
- (124) *Lehtiartikkeli **heräsi** (pro **herätti**) huomiota ministeriöllä, vaikka siellä suotiin Papier und Druckin joulumarkkinat etukäteen.* (SA0098c)

Väärän verbijohdoksen valinta. Ryhmän toinen, pienempi keskittymä koostuu tapauksista, joissa verbi ja sen frekventatiivijohdos (ks. esim. VISK § 353, § 354, § 1735) on

sekoitettu keskenään. Tämä sekoittuminen aiheuttaa osassa tapauksista hienoisen mutta huomattavan merkitys- tai sävyeron, kuten esimerkeissä (125–129).

- (125) *Siellä oli jotain nyrkkitappelu ja siitä seurauksena Jungnerin sormi joutui Kiviniemen suuhun ja nykyinen pääministeri **pureskeli** (pro **puraisti**) vastakkais-
taan.* (PU0001)
- (126) *Kielentutkimus alkoi **tarkastaa** (pro **tarkastella**) suomea vähemmistökieltenä
eri puolilla maailmaa.* (UN0013b)
- (127) *Tiedemiesten arvioiden mukaan maapallon nykyinen keskilämpötila olisi
0,7°C alhaisempi, jos ihmisen toiminnan vaikutusta luontoon voisi **vältellä**
(pro **välttää**).* (VE0148)
- (128) *Sekä Madridissa että Sevillassa minun oli **opiskeltava** (pro **opittava**), että päi-
vän kuumuuden aikana nukutaan ja yöllä oleskellaan.* (SA0107g)
- (129) *Koulun jälkeen Lada **oppi** (pro **opiskeli**) huonemaalariksi, mutta työ oli hä-
nelle liian raskas, joten hän vaihtoi alaa ja hänestä tuli kirjansitoja-kultaaaja.*
(TS0014l)

Esimerkeissä (130) ja (131) verbin ja sen frekventatiivijohdoksen sekoittaminen keske-
nään aiheuttaa puolestaan suuren ja ratkaisevan merkityseron, sillä verbiparin vaihtoeh-
dot tarkoittavat jo selvästi eri asiaa, vaikka ne onkin johdettu samasta kantasana-

- (130) *Päätettyäni siitä, mitä asioita tekstissäni **käsitän** (pro **käsittelen**), yritän heti
muodostaa ensimmäisen lauseen, jonka henkilökohtaisesti pidän tekstin ko-
vimpana osana.* (PU0027f)
- (131) *Se ei ole kaikkea, mitä tapahtui tänään, mutta tarvitsen unta niin kovasti, että
siitä hirvittävästä tapahtumasta, joka minun piti **kokeilla** (pro **kokea**), kirjoi-
tan vasta huomenna.* (PU0027g)

4.3. Morfologiset epätyypillisyydet

Morfologisten epätyypillisyyksien pääryhmä on oma lisäykseni Kuuluvaisen (2015) al-
kuperäisiin pääryhmiin, ja se koostuu tapauksista, joiden epätyypillisuus ilmenee sanan
muodossa. Käytännössä tämä tarkoittaa sitä, että joko suomenoppija on taivuttanut sanan
sellaiseen muotoon, jossa se ei tyypillisesti esiinny, tai sitten hän on valinnut sanalle sel-
laisen sijapäänteen, joka ei sovi sille virkkeen kontekstissa. Osa epätyypillisyyksistä on
sellaisia rajatapauksia, joita natiivikin saattaa käyttää, kun taas osa kulkee jo

epätyypillisyyden ja virheen rajamailla. Lisäksi on tapauksia, joissa suomenoppijan valitseman sananmuodon merkitys on niin kaukana virkkeen alkuperäisestä merkityksestä, että tapaus voitaneen jo laskea virheeksi. Ryhmä jakautuu kolmeen keskittymään, jotka ovat sanan epätyypillinen luku ja sijapäätte sekä *sti*-adverbijohdosten ylimuodostus.

Sanan epätyypillinen luku. Esimerkin (132) tapaus, jossa *työ* esiintyy yksikön illatiivissa, ei ole sinänsä harvinainen saati virheellinen. Tämän vahvistaa hakusanan TYÖ (YKS ILL) konkordanssitulokset, joissa osumia oli lähes 7 390. Silti *työ*-sanon käyttö monikon illatiivissa lieene huomattavasti yleisempää, sillä hakusana TYÖ (MON ILL) antoi lähes kaksin verroin osumia, vajaat 14 100.

- (132) *Seuraavana päivänä kun Mildred ja Lennox-Brown tulevat **työhön** (pro **töihin**), he huomaavat ikkunan takana oudon hiekkasäkin.* (PU0005r)

Esimerkissä (133) sanaa *väki* on käytetty monikossa. Tällainen käyttö on kuitenkin hyvin epätyypillistä, sillä osumia hakusanalla VÄKI (YKS) oli reilut 20 200, kun taas hakusana VÄKI (MON) antoi vain hieman yli 130 osumaa, joista osa on *väistää*-verbin kieltomuotoja tai yksikön toisen persoonan imperatiivimuotoja ja osa taas *välillä*-sanoja, joissa on lyöntivirhe. Lisäksi ne osumat, jotka oikeasti sisälsivät *väki*-sanon, olivat ylivoimaisesti muodossa *väke-* + OMISTUSLIITE.

- (133) *Hän tuli saarnaajaksi ja rupesi matkustamaan koko Mekongin suiston ympäri selittäen Jumalan tahdon **väille** (pro **ihmisille**).* (PU0013j)

Esimerkissä (134) ilmausta *korvan juuressa* on käytetty epätyypillisesti monikossa. Tapauksen epätyypillisyyden vahvistaa se, että hakuyhdistelmä *korvien + juurissa* ei antanut yhtään osumaa, kun taas osumia hakuyhdistelmällä *korvan + juuressa* oli reilu kymmenen. Myös esimerkin (135) *kaikesta huolimatta* -ilmauksessa sana *kaikki* on epätyypillisesti monikon muodossa. Hakuyhdistelmä *kaikista + huolimatta* antoi yhden ainoan osuman, kun taas osumia hakuyhdistelmällä *kaikesta + huolimatta* oli reilut 920.

- (134) *Rangaistukseksi J. A. ammuttaa revolverillaan veljesten **korvien juurissa** (pro **korvan juuressa**).* (PU0022h)

- (135) *Pitää rakastaa omaa puolisoaan **kaikista** (pro **kaikesta**) huolimatta, koska jokainen voi erehtyä, mutta riitoja voi karttaa olemalla pitkämielinen ja kärsivällinen.* (VE0159)

Sanan epätyypillinen sijapäätte. Esimerkin (136) ilmaus *auringon säteissä* on epätyypillinen siinä mielessä, että sana *säde* esiintyy vain harvoin inessiivissä. Hakusanojen SÄDE ja SÄDE (INE) konkordanssitulokset paljastavat, että SÄDE-sanan yli 1 580 osumasta vain viisi on inessiivissä. Sen sijaan hakusanojen PAISTE ja PAISTE (INE) konkordanssitulokset osoittavat, että PAISTE-sanan vajaan 250 osumasta lähes 80 on inessiivissä. Lisäksi ne ovat kaikki yksikössä, ja suurin osa viittaa nimenomaan auringon paisteeseen.

(136) *Siitä, että kuisti on auringon säteissä (pro paisteessa), voidaan päätellä, että on aamu.* (VE0165a)

Esimerkkejä (137) ja (138) voi luultavasti pitää jo virheinä, sillä eri sijapäätteen käyttö niissä johtaa ratkaisevaan merkityseroon. Esimerkin (137) *maa* on taivutettu ulkopaikallissijassa, mikä antaa sille merkityksen 'maaseudulla' (KS s.v. *maa*). Jos halutaan välittää merkitys, jossa lääketieteen opiskelija makaa maan pinnalla, on siis käytettävä sisäpaikallissijaa. Myös esimerkin (138) *väli*-sanon käyttö sisäpaikallissijassa ei sovi virkkeen kontekstiin, sillä se viittaa lähinnä kahden konkreettisen esineen väliseen tilaan. Sen sijaan sanan käyttö ulkopaikallissijassa viittaisi enimmäkseen abstraktiin etäisyyteen tai osapuolten keskinäiseen suhteeseen, kuten esimerkissä on ilmeisesti haluttu ilmaista (KS s.v. *välissä, välillä*).

(137) *Lääkisopiskelija on maalla (pro maassa), Kirsi osoittaa, että Lääkisopiskelija tarvitsee apua.* (SA0086c)

(138) *Työnjako lehtien välissä (pro välillä) on nyt erilainen.* (SA0097d)

sti-adverbijohdosten ylimuodostus. Kolmannen keskittymän tapauksissa *sti*-johtimella muodostettua adverbia on käytetty tilanteissa, joihin muun muassa genetiivimäärte (ks. VISK § 621, § 677) sopisi paremmin. *sti*-johdosten suosio lienee ymmärrettävä, sillä ne ovat adverbijohdostyypeistä produktiivisin ja yleisin (VISK § 373). Aivan kaikista adjektiiveista *sti*-johdosta ei kuitenkaan voi muodostaa (VISK § 62, § 373). Lisäksi *sti*-johdoksiin liittyy tiettyjä käyttöpreferenssejä, joista suomenoppija ei ole aina tietoinen.

Esimerkeissä (139) ja (140) *sti*-adverbia on käytetty toisen adverbien määritteenä tavan tai intensiteetin asteen ilmaisemiseksi. Esimerkin (139) tapaus ei ole sinänsä erityisen epätyypillinen, sillä hakuyhdistelmän SUHTEELLISESTI + ADV yli 700 osumasta reilu

kymmenen oli *suhteellisesti paljon*. Kuitenkin genetiivimuotoinen *suhteellisen* näyttää olevan huomattavasti yleisemmin käytössä adverbien yhteydessä, sillä hakuyhdistelmä *suhteellisen* + ADV antoi yli 1 650 osumaa, josta reilut 130 oli *suhteellisen paljon*. Esimerkin (140) epätyypillisuus on taas selkeämmin nähtävissä, sillä hakuyhdistelmän RIITÄVÄSTI + ADV yli 330 osumasta yksikään ei ollut *riittävästi paljon*. Sen sijaan hakuyhdistelmä *riittävän* + ADV antoi yli 2 750 osumaa, josta lähes 90 oli juuri *riittävän paljon*.

(139) *Hämmästykseni ymmärrän suhteellisesti (pro suhteellisen) paljon.* (PU0023c)

(140) *Hän ei kuitenkaan tee riittävästi (pro riittävän) paljon työtä, koska paljon aikaa menee hukkaan hänen viettäessään päiviä Siskon kanssa ajattelematta esitettä.* (PU0027l)

Esimerkeissä (141) ja (142) *sti*-adverbia on käytetty lauseen adjektiivipredikatiivin määrittämiseen. Kuitenkin esimerkin (141) *hirveä* sopii kyseiseen tarkoitukseen ainoastaan genetiivimuotoisena, ei *sti*-johdoksena. Tämä näkyy hakuyhdistelmän HIRVEÄSTI + ADJ yli 220 osumasta, joissa *hirveä*-sanana määrittämä adjektiivi ei esiinny predikatiivin roolissa eikä ilmausta HIRVEÄSTI KALLIS ole ollenkaan. Sen sijaan hakuyhdistelmä *hirveän* + ADJ antoi yli 2 870 osumaa, joista 25 on muodossa *hirveän KALLIS*. Myös esimerkin (142) *yllättävä* ei sovi *sti*-johdoksena adjektiivipredikatiivin määritteeksi. Osumia hakuyhdistelmällä YLLÄTTÄVÄSTI + ADJ oli vain vajaat 20, kun taas hakuyhdistelmä *yllättävän* + ADJ antoi lähes 3 450 osumaa, josta yksi on muodossa *yllättävän ILOINEN*.

(141) *Tämä on hirveästi (pro hirveän) kallista, siksi moni mieluummin maksaa, kuin pitää kiinni oikeuksistaan.* (SA0097)

(142) *Kuusamolaiset ovat yllättävästi (pro yllättävän) iloisia ja näyttävät onnelliselta.* (PU0027g)

Adverbien ja adjektiivien määritteen lisäksi *sti*-johdosta on käytetty myös tilanteissa, joissa tyypillisempi vaihtoehto olisi adverbiksi kivetynyt infiniittinen muoto (ks. VISK § 390). Vaikka esimerkin (143) *sti*-johdosta käytetäänkin jossain määrin natiivisuomessa, E-infinitiivin instruktiivi *yllättäen* on huomattavasti yleisempi, sillä osumia oli vajaat 240 hakusanalla YLLÄTTÄVÄSTI mutta yli 8 670 hakusanalla *yllättäen*. Myös esimerkin (144) *odottamattomasti* on harvemmin käytössä kuin MA-infinitiivin abessiivi *odottamatta*,

sillä hakusanalla ODOTTAMATTOMASTI osumia oli reilu 30 mutta hakusanalla *odottamatta* niitä oli melkein 150.

- (143) *Kuitenkin vertaamalla kielten typologisia etäisyyksiä, eli määrää kielten yhteisiä piirteitä, Dahl näyttää, että tällä ehdolla suomen lähimmät kielet ovat yllättävästi (pro yllättäen) enimmäkseen indoeurooppalaisia.* (PU0027n)
- (144) *Naiset eivät edes haaveile, että hän voisi tulla käymään, mutta hän kuitenkin odottamattomasti (pro odottamatta) tulee.* (PU0002h)

Esimerkkien (145–147) *sti*-johdokset ovat natiivisuomessa joko hyvin harvoin käytössä tai niitä ei muodosteta juuri lainkaan. Esimerkin (145) *todellisesti* on toki lueteltu Kieli-toimiston sanakirjassa, joskin siinä myös mainitaan, että tavallisempi muoto on *todella* (KS s.v. *todellisesti*). Lisäksi hakusana TODELLISESTI antoi vain reilut 30 osumaa, kun taas osumia oli TODELLA-hakusanalla peräti yli 71 000.

- (145) *Ja tarvitaan todellisesti (pro todella) suurta tapahtumaa, jotta se sytytettäisiin elämään, olipa se itsenäisyyden menetys, urheilujoukkueen voitto tai jonkun tärkeän henkilön kuolema.* (PU0005n)

Esimerkin (146) *sellaisesti* on epätyypillinen tai jopa virheellinen, sillä se kuuluu niihin *sti*-adverbeihin, joiden johtamista on rajoitettu (VISK § 373). Myös esimerkin (147) *nykyisesti* ei ole juurikaan käytössä. Tapausten epätyypillisuus varmistuu hakusanojen SELLAISESTI ja NYKYISESTI konkordanssituloksista, joissa kumpikaan hakusana ei antanut yhtään osumaa.

- (146) *Vahinko, että ihmiset eivät halua puhua sellaisesti (pro sellaisella tavalla) ja kieli muuttuu yksinkertaisemmaksi.* (PU0015f)
- (147) *Saksassa nykyisesti (pro nykyään) keskustellaan osavaltioiden kulttuuriministeriöiden ja kustantamoiden ehdottamasta uudesta ohjelmasta, jota pitää asentaa kouluissa käytettyihin tietokoneisiin.* (SA0097c)

4.4. Lauseketason epätyypillisyydet

Lauseketason epätyypillisyyksien pääryhmä koostuu tapauksista, joissa epätyypillisuus ilmenee yksittäistä sanaa laajemmassa mutta kokonaista lausetta suppeammassa tekstijaksossa. Se on rinnastettavissa Kuuluvaisen (2015) luokittelun *virheet ilmaisun*

rakenteessa -pääryhmään, jonka hän jakaa *lausekevirheiden* ja *yksittäisten ilmaisukeskittymien* alaryhmiin. Jälkimmäisessä alaryhmässä hän erottaa vielä kolme keskittymää, jotka ovat 1) virheet verbin ja sen täydennyksen semanttisessa kokonaisuudessa, 2) *kanssa*-sanana epätyypillinen käyttö sekä 3) sijapäänteen korvaaminen adpositiolla. Näiden lisäksi hän sisällyttää samaan ryhmään adverbiaalin käyttöön liittyviä virheitä.

Olen päätenyt omassa luokittelussani jakamaan ilmaisun rakenteen epätyypillisyydet lauseke- ja lausetasolle sillä perusteella, esiintyykö epätyypillisyyden kokonaisuudessa finiittiverbi vai ei. Tämän pääryhmän neljä alaryhmää ovat **epätyypilliset substantiivilausekkeet, adjektiivi- ja partisiippilausekkeet, adpositiolausekkeet** sekä **infinitiivilausekkeet**. Kuuluvaisen mainitsema virheet verbissä ja sen täydennyksissä siirtyvät siis lausetason epätyypillisyyksien pääryhmään, ja muut hänen mainitsemansa keskittymät sijoittuvat eri alaryhmiin sen mukaan, millaisessa lausekkeessa epätyypillisuus ilmenee.

Kuten mainitsinkin aiemmin, raja lauseketason epätyypillisyyksien ja useampisanaisten uudismuodosteiden (ks. luku 4.2.5.) välillä ei ole aina selkeä tai yksiselitteinen. Olen asettanut erottavaksi tekijäksi sen, että siinä missä useampisanaiset uudismuodosteet voi yleensä korvata yhdellä sanalla, lauseketason epätyypillisyyksien korvaavat vaihtoehdot ovat yleensä nekin useampisanaisia. Tämä johtuu siitä, että tyypillisemmän vaihtoehdon muodostaminen ei vaadi muutosta ainoastaan lausekkeen edussanassa vaan myös sen laajennuksissa (ks. VISK § 439). Tällöin useampi lausekkeen osa on muutettava.

4.4.1. Substantiivilausekkeet

Substantiivilausekkeiden alaryhmä koostuu tapauksista, joiden epätyypillisuus ilmenee useammassa substantiivilausekkeen osassa. Käytännössä epätyypillisuus voi koskea joko lausekkeen edussanaa eli substantiivia sekä sen laajennuksia tai vain osaa niistä, tai vaihtoehtoisesti se voi koskea vain useampia edussanan laajennuksista (ks. VISK § 442). Ryhmän kolme keskittymää sijoittuu jatkumolle, jonka yhdessä ääripäässä ovat pieniä muutoksia vaativat tapaukset ja toisessa ääripäässä taas ne tapaukset, jotka vaativat koko lausekkeen korvaamisen uudella rakenteella.

Sananvalinnan ja sananmuodon muutoksia vaativat epätyypillisyydet. Tämän keskittymän tapauksissa sekä epätyypillisemmän että tyypillisemmän substantiivilausekkeen osien määrä pysyy samana. Tarvittavat muutokset ovat suhteellisen pieniä ja koskevat esimerkiksi sananvalintaa tai taivutusta. Esimerkeissä (148) ja (149) on käytetty yksikkömuotoista ilmausta *hyvä yhteys* viittaamaan ihmisten keskinäisiin suhteisiin. Vakiintunut ilmaus on kuitenkin monikkomuotoinen *hyvät välit* (KS s.v. *väli*). Esimerkissä (150) käytetty *vakuutus* ja sen korvattavissa oleva *vakuus* muistuttavat muodoltaan toisiaan, mutta mikäli suomenoppija viittaa kirjoituksessaan 'varmuuteen jonkin asian paikkansapitävyydessä' (KS s.v. *vakuus*), jälkimmäinen sana sopii paremmin ja vaatii myös muutoksen *apu*-sanana taivutuksessa.

- (148) *He syövät joka päivä paistettuja kanoja, heillä ei ole **hyvää yhteyttä** (pro **hyviä välejä**) toisiinsa, vanhemmilla ei ole aikaa keskustella lastensa kanssa.* (PU0007d)
- (149) *Työkulttuuriin on oltava tutustua, jos ihminen haluaa menestyä vieraassa maassa ja olla **hyvässä yhteydessä** (pro **hyvissä väleissä**) työkavereiden kanssa.* (VE0093g)
- (150) *Kun Juhana sai **avun vakuutuksen** (pro **vakuuden avusta**) hän keräsi armeijaa ja syösi Eerikin valtaistuimelta.* (PU0001e)

Lausekkeen rakenteen muutoksia vaativat epätyypillisyydet. Tähän keskittymään kuuluvat tapaukset, joiden tyypillisemmän vaihtoehdon muodostaminen vaatii hieman isompia muutoksia substantiivilausekkeen osissa. Käytännössä tämä tarkoittaa muutoksia lausekkeen osien määrässä kieliopillisten ominaisuuksien lisäksi.

Esimerkin (151) vaihtoehtojen välillä on hienoinen mutta ratkaiseva merkitysero, sillä *minimaalinen määrä* välittää hieman eri merkityksen kuin sitä tyypillisempi vaihtoehto *minimimäärä*. Konkordanssihakua yhdistelmällä MINIMAALINEN & MÄÄRÄ antoi alle 20 osumaa, joiden mukaan ilmausta *minimaalinen määrä* käytetään lähinnä merkityksessä 'hyvin pieni määrä'. Sikäli kun suomenoppija on tarkoittanut kirjoituksessaan 'pienintä mahdollista määrää', sopivampi vaihtoehto olisi *minimimäärä*, sillä sanaa käytetään juuri tässä merkityksessä hakusanan MINIMIMÄÄRÄ yli 130 osumassa.

- (151) *Toinen ongelman ratkaisu on vähentää **minimaalista opiskelijoiden määrää** (pro **opiskelijoiden minimimäärä**) venäjän kieltä opiskelevissa ryhmissä.* (VE0130e)

Merkittäviä syntaktisia muutoksia vaativat epätyypillisyydet. Tämän keskittymän tapauksissa tyypillisemmän vaihtoehdon muodostaminen edellyttää koko substantiivilausekkeen korvaamisen jollakin toisella rakenteella. Esimerkeissä (152) ja (153) substantiivin etumääritteenä on käytetty partisiippilauseketta, joka vastaa funktioltaan substantiivia määrittävää relatiivilauseetta (ks. VISK § 531). Vaikka esimerkit ovatkin kieliopillisesti virheettömiä, esimerkin (152) *pakollisesti* ja esimerkin (153) *paikoillaan pysymätön* ovat suhteellisen harvoin käytössä verrattuna ilmauksiin *on pakko* ja *ei pysy paikoillaan*. Luontevampaa olisi siis käyttää relatiivilauseetta partisiippilausekkeen tilalla.

(152) ”*Pakollisesti katsottava elokuva* (pro *elokuva, joka on pakko katsoa*).”

(153) *Hän oli paikoillaan pysymätön lapsi* (pro *lapsi, joka ei pysynyt paikoillaan*), mutta hyvin herttainen ja hauska lapsi, että kaikki antavat anteeksi hänelle kuljeistaan. (VE0002d)

4.4.2. Adjektiivi- ja partisiippilausekkeet

Sisällytän tähän alaryhmään sekä **adjektiivilausekkeet** että adjektiivien tavoin käyttäytyvät **partisiippilausekkeet** (ks. VISK § 521, § 603, § 611), jotka toimivat substantiivia luonnehtivina kokonaisuuksina. Ryhmään kuuluvat tapaukset ovat enimmäkseen lauseen predikatiivin roolissa, ja niiden epätyypillisuus ilmenee sekä lausekkeen edussanassa eli adjektiivissa tai partisiipissa että sen laajennuksessa tai laajennuksissa.

Epätyypilliset adjektiivilausekkeet. Esimerkissä (154) on käytetty genetiivimääritettä *suunnattoman* komparatiivin *rikkaampi* määrittämiseen. Komparaatiojohdosta ei kuitenkaan voi määrittää tällaisella intensiteettimääritteellä perusasteisten adjektiivien tavoin, koska siihen sisältyy jo modifioiva johdinaines. Jotta ilmaus olisi natiivisuomelle tyypillinen, komparatiivijohdoksen edellä on siis oltava kvanttoriadverbi *paljon*, joka toimii sen määritteenä. (VISK § 633.)

(154) *Minun puolan sanastoni on myös suunnattoman rikkaampi* (pro *suunnattoman paljon rikkaampi*) ja synonyymejä tavallisesti löytyy vähäisellä vaivalla. (PU0005i)

Epätyypilliset partisiippilausekkeet. Esimerkissä (155) ilmausta *kiinni kasvanut* on käytetty merkityksessä 'paikalleen jähmettynyt', vaikka sen todellinen merkitys on 'kasvanut tai liitetty johonkin' (KS s.v. *kiinni*). Vakiintunut vaihtoehto olisi siis kuvaannollinen ilmaus *nauliintunut* ~ *nauliutunut paikalleen* (KS s.v. *nauliutua*). Esimerkin (156) ilmaus *voimakkaasti liittynyt* ei ole sinänsä epätyypillinen, sillä konkordanssihaku yhdistelmällä VOIMAKKAASTI & LIITTYÄ antoi yli 120 osumaa, josta merkittävä osa oli juuri muodossa LIITTYÄ VOIMAKKAASTI. Epätyypillistä lienee puolestaan se, että liittyvä entiteetti ei ole abstrakti käsite vaan kaupunki, sillä tällaista käyttöä ei ilmennyt haun osuissa. NUT-partisiipin sijaan sopivampi vaihtoehto olisi esimerkiksi olosijainen tilan adverbiaali *vahvasti yhteydessä*.

(155) *Sen jälkeen Kasia kiirehti katsomaan mitä oli tapahtunut ja hän seiso i kuin kiinni kasvanut* (pro *nauliintunut* ~ *nauliutunut paikalleen*), koska keittiö näytti niin, kuin trombi olisi juuri kuljennut läpi. (PU0031)

(156) *Huhtikuussa 1812 Helsingistä tehtiin Suomen uusi pääkaupunki, koska väitettiin Turun olevan liian voimakkaasti liittynyt* (pro *vahvasti yhteydessä*) ruotsalaiseen kulttuuriin. (PU0007c)

Esimerkissä (157) TU-partisiipin *unohdettu* yhteydessä on yritetty muodostaa tekijää ilmaiseva agenttiadverbiaali (ks. VISK § 1327, § 995) *elämän puolesta*. Ilmaus on epätyypillinen kahdesta syystä: ensinnäkin sanaa *puolesta* ei käytetä tällaisiin tarkoituksiin suomessa (KS s.v. *puolesta*), ja lisäksi tällaista rakennetta käytetään tavallisimmin silloin, kun tekijä on instituutio tai yksilöimätön joukko (VISK mp.) eikä abstrakti tarkoite kuten elämä. Sopivampi partisiippilauseke esimerkin tarkoitukseen olisi agenttipartisiipin (ks. VISK § 521, § 525) sisältävä *elämän unohtama*.

(157) *Mies ei muista mitään, mutta silti hän rupeaa asumaan toisten sataman köyhien, elämän puolesta unohdettujen* (pro *elämän unohtamien*) ihmisten kanssa suurissa satamassa olevissa konteissa, joita rahasta vuokraa sataman vartija Anttila (Sakari Kuosmanen). (VI0380)

4.4.3. Adpositiolausekkeet

Adpositiolausekkeiden alaryhmään kuuluvat tapaukset, joissa adpositiota – yleensä postpositiota – on epätyypillisesti käytetty sijapäätteen tilalla. Vaikka ryhmän nimi antaa

ymmärtää, että epätyypillisuus ilmenisi ainoastaan adpositiolausekkeissa, sisällytän siihen myös ne tapaukset, joissa esiintyy päinvastainen ilmiö eli adposition epätyypillinen korvaaminen sijapääteellä. Suurin osa tapauksissa ovat kuitenkin juuri sellaisia, joissa sijapäätte on korvattu adpositiolla.

Adpositio sijapääteen tilalla. Tällaisia epätyypillisyyksiä esiintyy aineistossani eniten adpositioiden *varten* ja *kanssa* käytössä. Sana *varten* on tarkoitusta ilmaiseva postpositio, joka ei vastaa mitään tiettyä taivutusmuotoa (VISK § 692). Tämä näkyy selvästi esimerkeissä (158–162), joissa tyypillisemmät vaihtoehdot edustavat kaikki erilaisia muotteja. Tyypillisemmän vaihtoehdon muodostaminen vaatii siis tietoa siitä, miten ilmaus muotoillaan eri tilanteissa ja eri sanojen yhteydessä. Toisin sanoen se vaatii laajaa tietoa sanojen primingeistä (ks. luku 3.1.5.), jollaista suomenoppijalla ei välttämättä ole hallussaan.

- (158) *Itse asiassa hän käyttää naisen hyväksi omaa huvia varten* (pro *omaksi huvikseen ~ ilokseen*). (KI0012)
- (159) *Kustaa Vaasa halusi, että Helsigistä tuli kilpailija lahden toisella puolella sijaitsevaa Tallinnaa varten* (pro *lahden toisella puolella sijaitsevalle Tallinnalle*). (PU0004e)
- (160) *Tšekki on hyvin vaikea kieli ja se on myös vaikea kieli runojen ja laulujen kirjoittamista varten* (pro *kirjoittamiseen*). (TS0002e)
- (161) *Vasta nyt kuitenkin ymmärsin, että myöskin suomen kielen vieraana kielenä oppimisprosessesi minun varten* (pro *minusta ~ mielestäni*) on kiinnostava. (VE0173a)
- (162) *Luulen, että se [tunteiden näyttämättömyys] on hyvä politiikkaa varten* (pro *politiikolle*) - politiikka ja diplomatia vaativat kykyä reagoida ja toimia nopeasti, mutta samalla ei näytä tunteitaan. (VE0203c)

Esimerkkien (163–166) kaltainen *kanssa*-sanan epätyypillinen käyttö on mainittu myös Kuuluvaisen (2015: 43) työssä, jossa hän arvioi sanaa käytettävän joko jonkin sijamuodon korvaamiseen tai transferin vaikutuksesta. Postpositiona *kanssa*-sanaa käytetään joko ”ilmaisema[an] kahden tasaveroisesti rinnasteisen tekijän suhdetta” tai merkityksessä ’seurassa, parissa, ohella, mukana, myötä’. Arkikielessä sitä käytetään myös ”vierasvoittoisesti ilmaisema[an] kahden ei täysin rinnasteisen tekijän suhdetta”, mikä ei ole kuitenkaan suositeltavaa nyky-yleisikielen normien puitteissa. (KS s.v. *kanssa*.) Myös

varten-sanan tavoin *kanssa*-tapausten tyypillisemmän vaihtoehdon muodostaminen ei seuraa tiettyä muottia, vaan se edellyttää laajaa tietoa sanojen primingeista.

- (163) *Esimerkiksi kaunis uusklassinen Senaatintori **Tuomiokirkon kanssa** (pro **Tuomiokirkkoineen**) ovat heidän suunnitelmaa.* (PU0004e)
- (164) *Suomessa pidin siitä, että kaikki sarjat, TV-showit, jne esitetään vain **tekstien kanssa** (pro **tekstitettyinä**).* (TS0002h)
- (165) *Englanti on myös liian erilainen **äidinkieleni kanssa** (pro **äidinkieleeni verrattuna**), mutta joskus huomaan pistäväni prepositioita tai postpositioita päättäen venäjän kielestä.* (VE0172)
- (166) *Osaako hän jo ääntää sanoja ”**r-kirjaimen kanssa**” (pro **r-kirjaimisia sanoja**) hyvin?* (PU0011b)

Postpositioiden *varten* ja *kanssa* lisäksi aineistossani esiintyy myös yksittäisiä tapauksia, joissa muita adpositioita on käytetty sijapääteen tilalla. Esimerkissä (167) kausaalista postpositiota *takia* on käytetty illatiivin (ks. VISK § 1243, § 1256) tilalla, kun taas esimerkissä (168) päähenkilöiden ominaisuutta on ilmaistu *mukaan*-postpositiolla ablatiivin (ks. VISK § 1244) sijaan.

- (167) *Se juoksi loputtomasti metsien ja soiden kautta ja loukkasi jalkansa vuoren **terävien kallioiden takia** (pro **teräviin kallioihin**).* (PU0015h)
- (168) *Lakaisijan sanojen avulla kertoja korostaa sitä, että päähenkilöt ovat samannäköisiä **eleiden ja ilmeiden mukaan** (pro **eleiltään ja ilmeiltään**).* (VE0128)

Sijapäätte adposition tilalla. Edelliseen keskittymään verrattuna tällaisia epätyypillisiä esiintyy aineistossani harvemmin. Esimerkeissä (169) ja (170) paikallissijaista adverbialia on käytetty adposition tilalla sanojen *kunnioitus* ja *kiinnostus* yhteydessä. Vaikka tällainen paikallissijaisten adverbialien ja adpositiolausekkeiden rinnasteinen käyttö onkin mahdollista joissain yhteyksissä (VISK § 689), allatiivin käyttö *kunnioitus*-sanon yhteydessä ja elatiivin käyttö *kiinnostus*-sanon yhteydessä ei näytä olevan tyypillistä (KS s.v. *kunnioitus*, *kiinnostus*).

- (169) *Minulla on paljon kunnioitusta **ulkomaalaisille** (pro **ulkomaalaisia kohtaan**), jotka tulevat Puolaan osaamatta puolan kieltä ja ihmettelen aina, miten he pärjäävät siellä.* (PU0005b)

- (170) *Maisteri jää pappilaan ja vaikka Elli ei ilmoita kiinnostusta **maisterista** (pro **maisteria kohtaan**), tämä haluaa mennä naimisiin Ellin kanssa.* (SA0032a)

Paikallissijojen päätteiden lisäksi epätyypillisyyksiä esiintyy myös muiden sijapäätteiden käytössä. Esimerkissä (171) komitatiivisijaista adverbialia *toisine ihmisineen* on käytetty merkityksessä 'toisten ihmisten kanssa'. Tällainen 'kanssa'-merkitys ei kuitenkaan sisälly komitatiivin merkitykseen, sillä komitatiivi ilmaisee pikemminkin jotakin kokonaisuutta osineen, seuralaisineen tai hallussa olevine entiteetteineen; se ei siis viittaa erillisiin olioihin, jotka ovat tekemisissä toistensa kanssa (VISK § 1264).

- (171) *Hän joi kallista viiniä, söi hedelmiä monista maista, pasteijoita satakielten kielistä ja muita kalliita herkuja, mutta hän oli surullinen ja tyytymätön, koska hän halusi huvitella **toisine ihmisineen** (pro **toisten ihmisten kanssa**).* (PU0028b)

4.4.4. Infinitiivilausekkeet

Infinitiivilausekkeiden alaryhmä koostuu tapauksista, joissa epätyypillisuus ilmenee A-, E- tai MA-infinitiivilausekkeessa (ks. VISK § 492), ja ryhmän tapaukset jakautuvat myös näiden kolmen infinitiivin keskittymiin. Suurinta osaa ryhmän tapauksista yhdistää se, että tyypillisemmän vaihtoehdon muodostaminen edellyttää infinitiivilausekkeen korvaamista kokonaan jollakin toisella rakenteella.

Epätyypillisyydet A-infinitiivilausekkeissa. Ryhmän pienimmän keskittymän muodostavat A-infinitiivin epätyypillisyydet. Esimerkin (172) finaalarakenteen (ks. VISK § 513) käyttö on epätyypillistä tai jopa virheellistä, sillä omistusliitettä vaativaa rakennetta ei voi käyttää passiivinmuotoisen verbin yhteydessä. Tyypillisemmän tai oikean vaihtoehdon muodostaminen edellyttäisi infinitiivilausekkeen korvaamista translatiivisijaisella prediatiiviadverbiaalilla (ks. VISK § 978).

- (172) *Toinen lukemani artikkeli oli kirjoitettu **kunioittaakseen Tarja Koposen muistoa** (pro **Tarja Koposen muistoksi**).* (PU0001j)

Epätyypillisyydet E-infinitiivilausekkeissa. Ryhmän suurimman keskittymän muodostavat E-infinitiivin epätyypillisyydet, ja niitä ilmenee useimmiten instruktiivin muodossa.

Nämä adverbiaalimäärittteenä toimivat E-infinitiivin instruktiivimuodot jakautuvat tapauksiin, joiden tyypillisempi vaihtoehto olisi essiivisijainen adverbiaali, sekä tapauksiin, joiden tyypillisempi vaihtoehto olisi toinen infinitiivityyppi.

Esimerkissä (173) ilmausta *luonnosta riippuen* on käytetty merkityksessä 'riippuvaisena luonnosta'. Sanalla *riippuen* on kuitenkin jo vakiintunut merkitys 'määräytyy jnk tekijän mukaan' (KS s.v. *riippuen*), joten ilmauksen käyttö ei sovi suomenoppijan tarkoittamaan kontekstiin. Esimerkeissä (174) ja (175) ilmauksia *laskien yhteen* ja *tehden yhteenvedon* on käytetty merkityksessä 'yhteenvetona'. Vaikka ilmaukset ovatkin kieliopillisesti virheettömiä ja semanttisesti ymmärrettäviä, konkordanssihaku yhdistelmällä *laskien + yhteen* ja *tehden + YHTEENVETO* ei antanut yhtään osumaa, kun taas osumia hakusanalla *yhteenvetona* oli reilut 20.

(173) *Jännää, sillä Puolassakin on aina asunut paljon ihmisiä, varsinkin maanviljelijöitä, jotka ovat eläneet koko elämänsä lähellä luontoa, jopa **luonnosta riippuen** (pro **riippuvaisena luonnosta**). (PU0022g)*

(174) ***Laskien yhteen** (pro **yhteenvetona**), suomettuminen on oikein monimutkainen ajatus. (PU0024a)*

(175) ***Tehden yhteenvedon** (pro **yhteenvetona**) minun täytyy myöntää, etten ole oikein ahkera kirjoittajana enkä kirjoita paljon, erityisesti kun ei ole pakkoa. (PU0027f)*

Esimerkissä (176) E-infinitiivin instruktiivimuotoa on käytetty ilmaisemaan päämäärään pääsemisen keinoa. Tällaiseen välinettä ilmaisevaan tarkoitukseen sopisi kuitenkin paremmin MA-infinitiivin adessiivimuoto (VISK § 517).

(176) *Me voimme parantaa tätä tietoisuutta **kehittäen** (pro **kehittämällä**) harrastuksiamme ja **keskittäen** (pro **keskittymällä**) elämässämme asioihin, jotka osaamme hyvin. (PU0027i)*

Epätyypillisyydet MA-infinitiivilausekkeissa. Sisällytän tähän ryhmään poikkeuksellisesti finiittimuotoiset tapaukset, sillä epätyypillisyyden ”keskipiste” on kuitenkin lauseen MA-infinitiivissä. Esimerkeissä (177) ja (178) MA-infinitiivin abessiivi on liitetty *olla*-verbiin, jonka kanssa se muodostaa muutos- tai tilapassiivin kielteisenä vastineena toimivan verbiliiton (VISK § 1341). Molemmissa tapauksissa *olla*-verbin ja transitiivisen MA-infinitiivin subjekti ovat samat. Tällöin tekijä on ilmeinen ja rakenteeseen sisältyy

implisiittinen tarkoituksellisuuden merkitys, jota suomenoppija ei ole välttämättä aikonut kirjoituksessaan välittää. Tyypillisempi ja neutraalimpi ilmaisutapa olisi siis verbiliiton korvaaminen kieltomuodolla.

(177) *Päähenkilö tunnustaa Siskolle pelkäävänsä muutosta ja siitä muistuttavaa vettä, mutta uusi kaveri **on ymmärtämättä** (pro **ei ymmärrä**) tätä pelkoa.* (PU00271)

(178) *Romaanin hahmot **ovat käsittämättä** (pro **eivät käsitä**) sodan poliittista tarkoitusta eivätkä halua pyrkiä sodan käsitykseen.* (VE0101)

Esimerkin (179) *olla tekemässä* -rakenne on verbiliitto, joka ilmaisee muun muassa käynnissä olevaa tapahtumaa (VISK § 452). Lisäksi rakenne kuvaa tyypillisesti duratiivista mutta tilapäistä tapahtumaa, ja siihen sisältyy usein tarkoituksellisuuden implikaatio (VISK § 1519). Esimerkin ilmaus *olin itkemässä* tuntuu oudolta, sillä rakenteen tilapäisyyden merkitys on ristiriidassa *koko päivän* -ilmauksen kanssa, ja lisäksi tuskin kukaan itkee tarkoituksella. Samoin kuin edellisissä esimerkeissä tyypillisempi ja luontevampi ilmaustapa olisi synteettisen verbimuodon käyttö verbiliiton sijaan.

(179) *Kaikki sujui ihan hyvin, mutta heidän mentyään pois minä **olin itkemässä** (pro **itkin**) koko päivän.* (VE0129f)

4.5. Lausetason epätyypillisyydet

Lausetason epätyypillisyyksien pääryhmä koostuu tapauksista, joissa epätyypillisuus ilmenee lauseessa eli sanojen ja lausekkeiden rakenteellisessa kokonaisuudessa, jonka ytimenä on finiittiverbi eli verbin persoonamuoto (VISK § 864). Ryhmä jakautuu kuuteen alaryhmään, jotka ovat **verbin valenssi, verbi ja sen täydennysympäristö, lausetyypit, tulla joksikin -rakenne, vierasperäiset lauserakenteet** sekä **kiteytyneet ilmaukset**.

Tämä pääryhmä vastaa *syntaktisten virheiden* pääryhmää Kuuluvaisen (2015) luokittelussa. Hänen erottamansa neljä virhetyyppikeskittymää siinä ovat 1) lausetyypin valintaan liittyvät virheet, 2) muodollisen subjektin käyttö, 3) virheet kysyvän sivulauseen muodostamisessa sekä 4) virheet tekstuaalisen kerrontamuottien muodostamisessa. Näistä lausetyypin valintaan liittyvät virheet olen säilyttänyt luokittelussani omana

alaryhmänä. Virheet kysyvän sivulauseen muodostamisessa sisältyvät puolestaan vierasperäisten lauserakenteiden alaryhmään, ja virheet tekstuaalisen kerrontamuottien muodostamisessa ovat rinnastettavissa kiteytyneiden ilmausten alaryhmään. Muodollisen subjektin käyttöä en käsittele lainkaan tässä tutkimuksessa.

4.5.1. Verbin valenssi

Verbin valenssi tarkoittaa sen kykyä saada oheensa täydennyksiä eli pakollisia seuralaisia lauseessa (VISK § 446). Kunkin verbin luontaiseen valenssiin kuuluu tietty määrä täydennyksiä, joiden pakollisuus johtuu esimerkiksi semanttisista syistä ja muodostaa jatkumon tiukasta lähes valinnaiseen (VISK § 447). **Verbin valenssin** alaryhmä koostuu siis tapauksista, joissa verbillä on eri valenssi kuin se, mitä siltä tyypillisesti odotetaan.

Esimerkissä (180) verbiä *ottaa* on käytetty tilanteessa, jossa sen valenssiin kuuluva objekti puuttuu. Tämän täydennyksen pois jättäminen saa lauseen kuulostamaan oudon keskeneräiseltä. Samoin *kaivata*-verbin valenssiin kuuluu yleensä kaksi täydennystä, kaipaaja ja kaipuun kohde. Esimerkissä (181) kaipuun kohde puuttuu, eli verbin valenssi kyseisessä käyttöyhteydessä on ristiriidassa sen tyypillisen valenssin kanssa.

(180) *Miehet haluavat ahdistaa sen [porsaan] johonkin nurkaan ja sitten **ottaa** (pro **ottaa sen kiinni**). (VE0125)*

(181) *Tiedän myös perheeni **kaipaavan** (pro **kaipaavan minua**), haluan heillekin tehdä pienen yllätyksen – ensimmäinen kotimatkani on siirtynyt marraskuun alusta lokakuun loppuksi.... (PU0022)*

Esimerkeissä (182) ja (183) verbiä *aiheuttaa* on käytetty lauseessa, josta sen valenssiin kuuluva objekti puuttuu. Tämän täydennyksen pakollisuus on tiukin juuri *aiheuttaa*-verbin kaltaisilla suhdetta ilmaisevilla verbeillä (VISK § 447), joten verbin käyttöä seuraavissa yhteyksissä voidaan pitää epätyypillisenä. Konkordanssihaku yhdistelmällä AIHEUTTAA + [0–1 sana] + ETTÄ paljastaa toki, että kyseinen rakenne on jossain määrin käytössä natiivisuomessakin, sillä osumia oli noin 50. Käyttö jäänee kuitenkin suhteellisen harvinaiseksi, sillä konkordanssihaku yhdistelmällä AIHEUTTAA + SE + [0–1 sana] + ETTÄ antoi moninkertaisen osumamäärän, vajaat 1 100.

- (182) *Nämä muutokset aiheuttivat* (pro *aiheuttivat sen ~ johtivat siihen*), *että kaikki maalaiset siirtyivät kolhooseihin ja sohvooseihin tai kapunkien alueelle.* (PU0006j)
- (183) *Ehkä voisit yksinkertaisesti aiheuttaa, että unohdan* (pro *saada minut unohtamaan*) *hänen nimensä.* (PU0011d)

Esimerkissä (184) käytetty *tutustua* on rektioverbi, joka vaatii illatiivimuotoisen täydennyksen (VISK § 487, § 1225). Tämä täydennys on siis lisättävä esimerkin lauseeseen, jotta se olisi natiivisuomelle tyypillinen.

- (184) *Jussi tutustuu myös* (pro *tutustuu myös siihen*), *kuinka kolhoosi-myymälät toimivat.* (PU0006j)

Myös *luottaa*-verbin valenssiin kuuluu kaksi täydennystä, luottaja ja luottamuksen kohde (VISK § 439, § 446). Jälkimmäinen täydennys puuttuu esimerkistä (185), minkä perusteella tapausta voitaneen pitää epätyypillisenä. Tapauksen epätyypillisuus ei ole kuitenkaan yhtä selvä ja yksiselitteinen kuin aiemmissa esimerkeissä, sillä konkordanssihaku yhdistelmällä LUOTTAA + [0–1 sana] + ETTÄ antoi yli 920 osumaa. Siitä huolimatta, että osumia hakuyhdistelmällä LUOTTAA + SIIHEN + [0–1 sana] + ETTÄ oli huomattavasti enemmän (melkein 2 000), konkordanssitulokset kertovat toisen vaihtoehdon olevan yleisesti käytössä natiivisuomessakin.

- (185) *Toisaalta saattamme luottaa* (pro *luottaa siihen*), *että sivustot kuten Facebook ovat ikuisia ja toisaalta olemme usein liian laiskoja tallentamaan jokapäiväisten tapahtumien muistoja jonnekin muualle, toiselle laitteelle.* (PU0004c)

4.5.2. Verbi ja sen täydennysympäristö

Verbin ja sen täydennysympäristön alaryhmä koostuu tapauksista, joissa epätyypillisuus ilmenee pääosin lauseen predikaatiksi valitun finiittiverbin ja sen täydennysympäristön (ks. VISK § 446) kokonaisuudessa. Esimerkissä (186) predikaatiksi valittu verbi *ansaita* sopii lauseen kontekstiin hyvin. Ilmaus *ansaitsee vaakunnassa olemista* on kuitenkin epätyypillinen muun muassa siitä syystä, että *ansaita*-verbin objektitäydennys on harvoin *minen*-johdos. Hakuyhdistelmä ANSAITA & -MINEN antoi yli 470 osumaa, joista

vain alle 20:ssä esiintyi *minen*-johdos *ansaita*-verbin objektina. Osuus on vielä tätäkin pienempi siihen nähden, että osumia ANSAITA-hakusanalla oli yhteensä yli 11 910.

- (186) *Ja he päättivät samalla hetkellä: Se on eläin joka ansaitsee vaakunassa olemista* (pro *ansaitsee paikan vaakunassa*)! (PU0032b)

Seuraavassa kahdessa esimerkissä epätyypillisuus koskee verbin täydennysympäristön lisäksi myös itse verbiä. Esimerkissä (187) ilmausta *kannattaa lisätä* on käytetty ennen *että*-sivulauseetta merkityksessä 'mainittakoon'. Tällaista käyttöä ei kuitenkaan näytä esiintyvän natiivisuomessa, sillä hakuyhdistelmän KANNATTAA + LISÄTÄ vajaasta sadasta osumasta yksikään ei viitannut siihen.

- (187) *Kannattaa liisätä* (pro *mainittakoon*), *että 1600-luvulla paitsi Turun yliopistoa, Suomessa oli myös olemassa monta muuta korkeakoulua*. (PU0032a)

Esimerkissä (188) sanaa *tappaminen* on käytetty *tarvita*-verbin objektitäydennyksenä relatiivilauseessa. *minen*-johdoksen käyttö *tarvita*-verbin objektina ei ole sinänsä epätyypillistä, sillä hakuyhdistelmä TARVITA + [0–1 sana] + -MINEN (OBJ) antoi 600 osumaa, joista suurimmassa osassa johdosta käytetään juuri tällä tavalla. Sanan *tappaminen* käyttö *tarvita*-verbin objektina on kuitenkin epätyypillistä, sillä se ei ilmennyt yhdessäkään hakuyhdistelmän TARVITA & TAPPAA (-MINEN, OBJ) yli kymmenessä osumassa.

- (188) *Kuitenkaan hän ei ole oikeudenmukaista kostajaa, koska hän on psykopaatti, joka tarvitsee tappamista* (pro *jonka täytyy päästä tappamaan*). (TS0002h)

4.5.3. Lausetyypit

Lausetyypien alaryhmä vastaa lausetyypin valintaan liittyvien virheiden keskittymää Kuuluvaisen (2015) luokittelussa. Se koostuu tapauksista, joissa suomenoppija on valinnut epätyypillisen syntaktisen lausetyypin sanomansa välittämiseksi. Ryhmän kaksi keskittymää ovat omistuslauseen ja nesessiivirakenteen epätyypillinen käyttö. Kolmas samantyyppinen ilmiö, *tulla joksikin* -rakenteen ylikäyttö esiintyy sen verran laajasti aineistossani, että omistan sille oman alalukunsa.

Omistuslauseen epätyypillinen käyttö. Tämä keskittymä koostuu tapauksista, joissa omistuslauseessa esiintyvä omistaja tai omistettava kohde on jollakin tavalla epätyypillinen. Esimerkeissä (189) ja (190) omistuslausetta on käytetty kontekstissa, jossa se ei välttämättä olisi semanttisesti sopivin valinta. Törmäykset ja avioliitot eivät nimittäin liene tavallisimpia asioita, joita pidetään omistettavissa olevina. Tyypillisempi lausetyyppi olisi vaikkapa intransitiivilause *hän joutui kolariin* esimerkille (189) ja kopulalause *isäpuoli on jo ollut avioliitossa ~ naimisissa* esimerkille (190).

(189) *Ehkä... jos **hänellä oli törmäys** (pro **hän joutui kolariin**), hän voisi soittaa veljelleen eli enolleni.* (PU0001d)

(190) *Tekstistä tulee ilmi, että **isäpuolella on jo ollut avioliitto** (pro **isäpuoli on jo ollut avioliitossa ~ naimisissa**).* (VE0121)

Esimerkissä (191) omistuslausetta on taas käytetty kontekstissa, jossa omistaja *ammatti* on eloton. Tämä on epätyypillistä, sillä omistaja on omistuslauseessa tyypillisesti elollinen (VISK § 895). Tyypillisempi ilmaus olisi esimerkiksi *ammatit, joissa pystytään auttamaan toisia ihmisiä*.

(191) *Minusta ammatit, **joilla on kyky auttaa** (pro **joissa pystytään auttamaan**) toisia ihmisiä, ovat aina olleet arvostetuimmat.* (SA0097)

Nesessiivirakenteen epätyypillinen käyttö. Kaikki tämän keskittymän tapaukset liittyvät *on tehtävä* -tyyppisiin, passiivin VA-partisiipin sisältäviin nesessiivirakenteisiin (ks. esim. VISK § 906, § 1580, § 1581, § 1670). Rakenteen käyttö esimerkeissä (192) ja (193) on epätyypillistä, sillä se ei tavallisesti esiinny kielteisessä yhteydessä. Kieltoverbin kanssa sopisi tyypillisemmin välttämättömyyttä ilmaiseva modaaliverbi *tarvita* (ks. esim. VISK § 1564, § 1613).

(192) *Varmasti **minun ei ole puhuttava** (pro **minun ei tarvitse puhua**) suomea osatakseni sämpylän suomalaista kaupasta, jos siellä niin kuin kaikkialla osataan englantia.* (PU0015f)

(193) ***Lukijan ei ole luettava** (pro **lukijan ei tarvitse lukea**) rivien välistä huomatakseni, että se pelko on nuoren miehen pakkomielle.* (PU0027d)

4.5.4. *Tulla joksikin* -rakenne

***Tulla joksikin* -rakenteen** alaryhmä liittyy lausetyyppien alaryhmään. Se koostuu tapauksista, joissa tilanmuutosmuottia *tulla joksikin* käytetään tilanteissa, joihin sen merkitys ei sovi. Yksi syy rakenteen ylikäyttöön voi olla se, että sitä tarjotaan ensimmäisinä hakutuloksina, kun sanakirjasta haetaan vastaavamerkityksistä verbiä toisella kielellä, esimerkiksi englannin kielen *become* (Sanakirja.org englanti–suomi s.v. *become*). Rakennetta käytetään epätyypillisesti useissa yhteyksissä yhtäältä tuloslauseen ja toisaalta refleksiivisen verbin sijasta. Lisäksi se esiintyy translatiivimuotoisen predikaatiiviadverbiaalin sisältävissä lauseissa, joihin sopisi jokin muu verbi kuin *tulla*.

Rakenteen käyttö tuloslauseen sijasta. Yleisimpiä aineistossani olivat tapaukset, joissa *tulla joksikin* -rakennetta käytetään tuloslauseen tavoin ilmaisemaan subjektia koskevaa muutosta. Lähimerkityksisyydestään huolimatta nämä kaksi rakennetta eivät ole täysin rinnastettavissa toisiinsa, sillä tuloslause ilmaisee pikemminkin tarkoitteen uuden ominaisuuden tai luokan syntyä, kun taas *tulla joksikin* -rakenteen ilmaisema muutos tapahtuu enimmäkseen jo olemassa olevassa ominaisuudessa tai luokassa. Näin ollen vain tuloslause sopii seuraavien esimerkkien kaltaisiin tapauksiin, joissa tarkoite kehkeytyy tietynlaiseksi mutta pysyy itse perusolemuksestaan samana. (VISK § 904.)

Tulla joksikin -rakenteen subjektin paikkaan on sijoitettu monenlaisia tarkoituksia. Yhtäältä subjektiksi on valittu elollinen tarkoite kuten esimerkeissä (194) ja (195) tai elollisten tarkoitteiden edustama entiteetti kuten esimerkeissä (196) ja (197). Toisaalta subjektina esiintyy myös eloton esine kuten esimerkissä (198) tai abstrakti käsite kuten esimerkeissä (199) ja (200). Tapaukset tekee epätyypilliseksi vielä se, että kyseessä on aikaa vievä toiminta, kuten esimerkeistä ilmenee, jolloin *tulla joksikin* -rakenne ei sovi vaan ainoa vaihtoehto on tuloslause (VISK § 904).

(194) *Vuonna 1946 Paasikivi tuli presidentiksi* (pro *Paasikivestä tuli presidentti*). (PU0006k)

(195) *Näin Nikita ja violetti apina tulivat ystäväksi* (pro *Nikitasta ja violetista apinasta tuli ystäviä*). (VE0002d)

- (196) *Kaikilla on visio, että **Brightdairy Oy** voi tulla maailman huippumaitobrändiksi* (pro ***Brightdairy Oy**:stä voi tulla maailman huippumaitobrändi*). (KI0003p)
- (197) *Todellisuudessa Suomalaiset saivat koko autonomiasa takaisin vasta vuonna 1917, kun **Suomi tuli itsenäiseksi maaksi*** (pro ***Suomesta tuli itsenäinen maa ~ Suomi itsenäistyi***). (PU0029a)
- (198) *Muuttaa sananmukaisesti, koska **metro on tullut hanelle kodiksi*** (pro ***metrostä on tullut hänelle koti***), jossa hän viettää päivät ja yöt. (PU0005f)
- (199) *Tulee vielä mainita, että **onnellinen yhteensattuma voi joskus tulla elämän käännekohtaksi*** (pro ***onnellisesta yhteensattumasta voi joskus tulla elämän käännekohta***), kun kaikki muuttuu aivan toisenlaiseksi. (VE0005a)
- (200) *Aurinko ja pilvet ovat kuvassani siksi, että kaikista hankalista puolistaan huolimatta **suomen kieli on tullut minulle hyvin läheiseksi*** (pro ***suomen kielestä on tullut minulle hyvin läheinen***) ja olen onnellinen, että osaan sitä niin hyvin. (VI0401a)

Suurin osa aineistossani esiintyvistä tapauksista koskee intransitiivisen tuloslauseen korvaamista sen translatiivimuotoisen predikatiiviadverbiaalin sisältävällä vastineella. Joitakin transitiivisia tapauksia kuitenkin esiintyy, kuten esimerkiksi (201).

- (201) *Vuonna 1556 Kustaa Vaasa **teki poikansa Juhanan Suomen herttuaksi*** (pro ***teki pojastaan Juhana Suomen herttuan***). (PU0022a)

Rakenteen käyttö refleksiivisen verbijohdoksen tilalla. Edellistä pienemmän keskittymän muodostavat tapaukset, joissa *tulla joksikin* -rakennetta käytetään refleksiivisen verbijohdoksen (ks. esim. VISK § 334, § 335) tilalla. Näissäkin tapauksissa muutos ei koske tarkoitteen perusolemusta, vaan muutoksen seurauksena tarkoitteelle kehittyi uusi ominaisuus. Tämän vuoksi rakenteen käyttöä esimerkeissä (202–204) voitaneen pitää epätyypillisenä.

- (202) *Sen vuoksi aika kuluu paljon ja ihmisten suhteet **tulevat jopa huonommiksi*** (pro jopa ***huononevat***). (KI0003q)
- (203) *Äiti yrittää rauhoittaa minua, mutta **en** minä **tule rauhalliseksi*** (pro ***en** [...] **rauhoitu***) ennen kuin tiedän, että Kaczyński kärsii musertavaa tappiota ja ettei hän enää hävittä ja nöyryyttää Puolaa. (PU0001d)
- (204) *Sitä minä tarvitsin, mutta usein sen jälkeen **tulin nolostuneeksi*** (pro ***nolostuin***) ja luulin sanoneeni liian paljon. (PU0005i)

***Tulla*-verbin käyttö translatiivimuotoisen predikatiiviadverbiaalin yhteydessä.**

Tämä keskittymä koostuu esimerkkien (205–207) kaltaisissa tapauksista, joihin sopii translatiivimuotoinen predikatiiviadverbiaali muttei verbi *tulla* sen yhteydessä. Myös näissä tapauksissa itse tarkoite pysyy aina samana, joten *tulla joksikin* -rakenne ei välttämättä sovi.

(205) *Kaiken lisäksi mies **tuli** (pro **joutui**) aivan käsittämättömän ryöstön uhriksi.*
(PU00221)

(206) *Tilanne **tulee** (pro **muuttuu**) vielä vaarallisemmaksi, kun Jussia syytetään Neuvostoliiton sotilaskarkuruudesta ja halutaan lähettää hänet armeijiaan.*
(PU0006j)

(207) *Kuitenkin, isännällä ei ollut aikaa ajatella sitä, koska hänen maatilansa **tuli** (pro **kasvoi**) koko ajan suuremmaksi ja täytyi hoitaa sitä.* (PU0027b)

Ryhmän viimeiseksi esimerkiksi jääköön seuraava muunnos kiteytyneestä kosintailmauksesta. Sen tyypillisyyden tai epätyypillisyyden jätän lukijan päätettäväksi.

(208) *Nähtyään koiran hän ilahtui ja kysyi: „Haluatko **tulla koirakseni**?”*
(PU0015h)

4.5.5. Vierasperäiset lauserakenteet

Vaikka vieraan kielen vaikutus onkin periaatteessa nähtävissä läpi koko aineistoni, olen koonnut **vierasperäisten lauserakenteiden** alaryhmään ne tapaukset, joissa lauserakenteen vierasperäisyys on erityisen ilmiselvää. Kyseessä ei siis ole yhden ”väärän” sanan tai muodon valinta sanakirjan hakutuloksista vaan laajemman rakenteellisen kokonaisuuden kääntäminen sanasta sanaan toisesta kielestä suomeksi.

Vaikka seuraavan kahden esimerkin epätyypillisyydet eivät yllä lauseen tasolle, sisällytän ne silti tähän ryhmään, sillä niiden tyypillisemmän vaihtoehdon muodostaminen vaatii muutoksen pikemminkin lauseen rakenteessa kuin sananvalinnassa. Kummassakin tapauksessa näkyy selvä viron kielen vaikutus. Esimerkin (209) ilmaus *ilman kysymättä* selittyy sillä, että toisin kuin suomessa viron abessiivia edeltää usein prepositio *ilma* (Eesti keele käsiraamat 2007: 252; EKI ühendsõnastik s.v. *ilma*). Esimerkin (210) *oppia*

selväksi kääntyy taas suoraan viron *selgeks õppida* -ilmauksesta, joka tarkoittaa jonkin asian oppimista kokonaan (EKSS s.v. *õppima*).

- (209) *En itsekkään ole pohdiskellut siitä, mutta nyt kuitenkin sain sille vastauksen, ilman* (pro *õ*) *kysymättä ja kovasti miettimättä.* (VI0399a)
- (210) *Koska itse opin suomen kielen selväksi* (pro *õ*) *jo ihan pienenä suomen koulussa ja ilman kaikenlaisia kieliopin miljoonaa sääntöä niin en oikeastaan tiedä miltä tuntuu oppia suomen kieltä "oikeasti".* (VI0399a)

Esimerkin (211) ilmaus – –, *jota voin ajatella* muistuttaa englannin ilmausta – – *that I can think of*. Samoin esimerkin (212) ilmaus *aika on ylös* muistuttaa englannin *time is up* -ilmausta. Esimerkki (213) edustaa puolestaan Kuuluvaisen (2015) virheet kysyvän sivulauseen muodostamisessa -keskittymää, joka kuuluu syntaktisten virheiden pääryhmään hänen luokittelussaan. Kysyvä sivulause – –, *jos Osku on syönyt* muistuttaa sekä englannin – – *if Osku has eaten* -vastinetta että Kuuluvaisen (mts. 46) mainitsemaa ruotsin *om du vill gå* -tyyppistä rakennetta. Vieraan kielen vaikutus on erittäin selvä myös esimerkeissä (214), sillä *auttaa*-verbiä ei käytetä alla olevassa yhteydessä suomessa lainkaan (KS s.v. *auttaa*). Ilmaus *en voi auttaa, mutta* muistuttaa vahvasti englannin *I can't help but* tai saksan *ich kann mir nicht helfen, [aber]* -rakennetta (MOT Sanakirjat saksa–suomi s.v. *helfen*), joka muodostetaan suomessa MA-infinitiivin abessiivin avulla.

- (211) *Ainoa digitaalivalokuvien huono puoli, jota voin ajatella* (pro *joka tulee mieleen*) *on se, ettemme nykyisesti voi olla varma, että valokuvassa esiinnytty ihminen tai maisema näyttää todellisuudessa samanlaista.* (PU0001c)
- (212) *Oli aika nopeasti ylös* (pro *Aika kului nopeasti*) *mutta kun tulin kotiin olin niin väsynyt että menin heti sänkyyn...* (RU0050c)
- (213) *Nyt Karvinen on menossa katsomaan jos Osku on* (pro *onko Osku*) *syönyt kulhosta.* (RU0019i)
- (214) *Vaikka kaikki oli hauska ja olen nähnyt niin paljon, en voi auttaa, mutta kaipa* (pro *en voi olla kaipaamatta*) *teitä ja Suomea hirveästi.* (SA0107g)

4.5.6. Kiteytyneet rakenteet

Kiteytyneiden rakenteiden alaryhmä koostuu kahdenlaisista tapauksista. Ensimmäisessä keskittymässä ovat tapaukset, joissa suomenoppija on selkeästi tavoitellut tiettyä

vakiintunutta rakennetta mutta epäonnistunut sen muodostamisessa. Toiseen keskittymään kuuluvat taas tapaukset, joissa suomenoppija on onnistunut muodostamaan tietyn vakiintuneen rakenteen mutta silti käyttänyt sitä epätyypillisellä tavalla.

Rakenteen epäonnistunut muodostaminen. Esimerkissä (215) suomenoppija on selvästi tavoitellut rakennetta *antaa jklle selkään* mutta valinnut sitä muodostaessaan sekä väärän verbin että väärän sijapäänteen habitiiviadverbiaalille (ks. VISK § 986).

(215) *Hän luuli Mikolaj-kokin **pehmittävän hänen selkään*** (pro *antavan hänelle selkään*). (PU0025b)

Esimerkeissä (216) ja (217) suomenoppija näyttää puolestaan hakeneen *mitä tulee johonkin* -rakennetta, joka esiintyy yleensä lauseen alussa. Esimerkissä (216) käytetty *mikä koskee jotakin* -rakenne näyttää tosin olevan jossain määrin käytössä natiivisuomessakin, sillä hakuyhdistelmä *Mikä + koskee* antoi kaksi tällaista osumaa. Silti *mitä tulee johonkin* -rakenne näyttää olevan paljon tyypillisempi, sillä hakuyhdistelmä *Mitä + tulee* antoi 270 osumaa, joista suurin osa sisältää juuri kyseisen rakenteen. Myös esimerkissä (217) rakenteen muodostaminen on epäonnistunut siinä, että siihen on valittu väärä verbi.

(216) *Mikä koskee kirjoittamista* (pro *mitä tulee kirjoittamiseen*), *olin samassa tilanteessa muiden kurssilaisten kanssa*. (VE0172)

(217) *Mitä kuuluu* (pro *tulee*) *minun ystäviini, he sanoivat vain "Kiitos, hei-hei!" ja sekunnin kuluttua katosivat näkyvistä*. (VE0129f)

Onnistuneesti muodostetun rakenteen epätyypillinen käyttö. Esimerkissä (218) ilmausta *miten voimme* on käytetty merkityksessä 'mitä meille kuuluu'. Ilmauksen käyttö ei välttämättä sovi kyseiseen kontekstiin, sillä suomenoppija näyttää viittaneen pikemminkin kuulumisiin kuin terveydentilaan (KS s.v. *voida*).

(218) *Lehdessämme lomalle emme tarvitse ottaa mukaan vihkojamme – jos paikalla on netti ja tietokone voimme aina kertoa **miten voimme*** (pro *mitä meille kuuluu*). (PU0001c)

Seuraavan kahden esimerkin epätyypillisuus jää puolestaan tulkinnanvaraisemmaksi. Esimerkissä (219) rakennetta *pitää jotakin t. jotakuta jonakin* on käytetty merkityksessä 'luulla joksikin t. jonkinlaiseksi', mikä onkin yksi sen merkityksistä (KS s.v. *pitää*),

joskin rakenteen yleisimmät käyttöyhteydet liittyvät nähdäkseni mielipiteeseen tai kannottoon. Esimerkissä (220) on käytetty permissiivirakennetta (ks. VISK § 502), jonka eloton subjekti *kyky* saattaa olla ristiriidassa rakenteen luvallisuuden implikaation kanssa. Kyseisten rakenteiden todelliset käyttöpreferenssit jäävät toistaiseksi epävarmoiksi, ja käsitkset esimerkkien epätyypillisyydestä saattavat vaihdella natiivienkin kesken.

(219) *Seuraavana aamuna Lennox **pitää** ikkunan takana olevaa Lammia **hiekkasäkinä** (pro **luulee** [...] **hiekkasäkiksi**) ja yrittää poistaa sen.* (PU0003k)

(220) *Koko sivulla kerrotaan kuvaaja Raimo Holapasta, jonka kyky huomata luonnon merkkejä **antaa hänen saada** (pro **auttaa häntä saamaan**) erittäin onnistuneita kuvia.* (PU0005j)

4.6. Yhteenveto ja tulosten merkitys oppijansuomen opetuksen kannalta

Olen esittänyt edellisissä alaluvuissa edustavan otoksen fraseologisista epätyypillisyyksistä, joita esiintyy ICLFIn C-tason oppijansuomen teksteissä. Olen luokitellut tapaukset neljään pääryhmään, jotka sijoittuvat eri kohtiin Kuuluvaisen (2015) kehittämällä leksikaalisuus–morfosyntaktisuus-jatkumolla. Epätyypillisyyksien paikkansapitävyyttä olen tarkastellut vertailuaineiston, sanakirjan sekä kieliopin avulla. Koska kieli muodostuu paljon laajemmasta kokonaisuudesta kuin tarkastelemistani osa-alueista, tulokseni ovat pikemminkin suuntaa antavia kuin ehdottomia.

Tulokseni edustavat sellaista edistyneen oppijansuomen muotoa, joka on ominaista ulkomaisissa yliopistoissa suomea vieraana kielenä opiskeleville. Osa epätyypillisyyksistä selittyy sillä, että näillä suomenoppijoilla on ollut vähemmän kosketusta kohdekieleensä kuin esimerkiksi Suomessa suomea toisena kielenä opiskelevilla tai omaksuvilla. Muun muassa Isohätälä (2014) on havainnut, että suomea toisena ja vieraana kielenä opiskelevien välillä esiintyy joitakin olennaisia eroja sanaston käytössä. Olisi siis mielenkiintoista saada selville, millaisia fraseologisia epätyypillisyyksiä esiintyy edistyneiden suomea toisena kielenä opiskelevien tuotoksissa verrattuna aineistoni epätyypillisyyksiin.

Esittelemäni tapaukset vaihtelevat epätyypillisyydeltään natiivisuomessakin esiintyvistä rajatapauksista varsinaisiin virheisiin. Ääripäiden välimaastoon mahtuu laaja kirjo tapauksia, jotka poikkeavat vaihtelevissa määrin natiivisuomesta. Sinclairin (1991)

näkemyistä (ks. luku 3.1.4.) noudattaen myös edistyneiden suomenoppijoiden tekstit rakentuvat pääosin idiomiperiaatteen mukaan. Aineistoni epätyypillisyydet kuitenkin osoittavat, että siirtyminen vapaan valinnan periaatteen varaan tapahtuu huomattavasti useammin oppijansuomessa kuin natiivisuomessa. Useimmissa aineistoni tapauksissa on myös nähtävissä Jantusen (2008) esittämiä oppijankielen universaaleja (ks. luku 3.3.1.), erityisesti kontekstuaalisen epäkonventionaalisuuden ja kieltenvälisen vaikutuksen ominaispiirteitä. Nämä poikkeavuudet kertovat, että suomenoppijoiden ja natiivien leksikaalissa ja morfologisessa primingeissa (ks. luku 3.1.5.) vallitsee selvä ero.

Epätyypillisyydet koostuvat sekä yksittäistapauksista että yleistettävissä olevista keskittymistä. Huomionarvoisia esiintymistiheydeltään ovat modaaliverbien ja intensiteettipartikkelien epätyypillinen käyttö sekä transitiivi- ja intransitiiviverbien sekoittaminen sananvalinnan pääryhmässä, *sti*-adverbijohdosten ylimuodostus morfologisten epätyypillisyyksien pääryhmässä, sijapäätteen korvaaminen adpositiolla lauseketason epätyypillisyyksien pääryhmässä sekä *tulla joksikin* -rakenteen ylikäyttö lausetason epätyypillisyyksien pääryhmässä.

Tulokseni vahvistavat muun muassa Nesselhaufin (2003: 237) ja N. Ellisin (2008: 7) näkemyksiä siitä, että natiivinkaltaisen idiomaattisen kielen tuottaminen saattaa olla hyvin vaikeaa jopa edistyneille kielenoppijoille. Lisäksi ne tukevat muun muassa Nesselhaufin (2005) ja Ivaskan (2015: 3–4) havaintoja siitä, että fraseologiset virheet tai epätyypillisyydet ovat tavallisia edistyneidenkin kielenoppijoiden tuotoksissa.

Tulokseni osoittavat myös, että edistyneen kielenoppijan todellinen sanastollinen ja kieliopillinen kompetenssi ei vielä täysin yllä Eurooppalaisen viitekehyksen (2003: 157–161) kyseiselle kielitaitotasolle asettamiin kriteereihin. Tutkimukseni suomenoppijat tekevät nimittäin vakaviakin sanastovirheitä erityisesti verbinvalinnassa, eivätkä he ole aina tietoisia valitsemiensa sanojen tyypillisistä käyttöympäristöistä ja primingeista. Vaikka kielenoppijoiden kuuluisi C-tasolla tuottaa EVK:n kriteereiden mukaan kieliopillisesti lähes virheetöntä kieltä, he tekevät todellisuudessa sellaisiakin virheitä, joiden voisi odottaa jääneen alkeis- tai keskitasolle (kielellisen kompetenssin taitotasoitteisista kriteereistä ks. esim. EVK 2003: 157–161; Opetushallitus 2003: 229–247).

Puutteellisesta fraseologisesta hallinnasta huolimatta tutkimieni edistyneiden suomenoppijoiden tuotokset ovat enimmäkseen kieliopillisesti johdonmukaisia ja semanttisesti ymmärrettäviä. Siinä missä suuri osa Kuuluvaisen (2015) analysoimista, pääosin A2–B1-tason tapauksista on merkitykseltään toisistaan kaukaisempia sanapareja, analysoimani C-tason tapaukset koostuvat enimmäkseen lähimerkityksisistä vaihtoehtoista.

Analysoimistani tapauksista voi myös havaita, että fraseologia on todellakin kollokaatiota paljon laajempi ilmiö, jonka hallinta vaatii runsaasti kosketusta autenttisiin kielenlähteisiin sekä laaja-alaista harjoitusta muun muassa sananvalinnan, morfologian sekä lauseke- ja lausetason rakenteiden parissa. Näin ollen tulokseni puoltavat sitä näkemystä, että kieltä tulisi aina opettaa yksittäisiä sanoja laajempina kokonaisuuksina. Tämä on tietenkin helpommin sanottu kuin tehty: vaikka kielenopetuksessa on viime aikoina alettu kiinnittää enemmän huomiota fraseologian tärkeyteen, sen sisällyttäminen käytännön opetukseen on vasta lapsenkengissä, ja sen opetus ulottuu vielä harvoin kollokaatioiden ja idiomien ulkopuolella oleviin rakenteisiin. (Lisää fraseologian opetuksen haasteista ks. esim. Kennedy 2008: 36–40; Meunier & Granger 2008: 247–249; Jantunen 2009b: 373–375; Brunni & Jantunen 2015: 391–392.)

Jantunen (2009b: 373–374) on aiemmin esittänyt kysymyksen siitä, kannattaako sanastoa ryhtyä opettamaan fraseologisina kokonaisuuksina jo kielenoppimisen alkumetreillä vai vasta myöhemmin edistyneemmällä tasolla. Hänen mukaansa tyypillisiä fraseologisia rakenteita pitäisi todellakin sisällyttää kielenopetuksen jo alusta alkaen. Tätä näkemystä puoltaa se, että tuloksissani esiintyy runsaasti ja laajasti sellaisiakin epätyypillisyyksiä tai virheitä, joita ei EVK:n (2003) kriteereiden mukaan kuuluisi C-tasolla enää olla. Sanojen käyttöyhteyksien opettaminen sanaston yhteydessä olisi kenties auttanut välttämään ainakin osan tutkimistani tapauksista.

Opetushaasteiden lisäksi Jantunen (mp.) on huomauttanut, että suomen kielen sanakirjat eivät toistaiseksi tarjoa riittävästi tietoa leksikaalisten yksiköiden fraseologisuudesta. Tämä näkyy tuloksissani erityisesti sanakirjan sudenkuopissa ja transfer-tapauksissa (ks. luku 4.2.6.), virheellisissä verbinvalinnoissa (ks. luku 4.2.8.) sekä vierasperäisissä rakenteissa (ks. luku 4.5.5.). On ilmiselvää, että sanoja on usein valittu sanakirjan hakutuloksista satunnaisin perustein ilman sen enempää tietoa niiden käyttökonteksteista. Yhdyn

siis Jantusen näkemykseen siitä, että opetuksen tueksi ei ole tällä hetkellä saatavilla riittävästi todelliseen kielenkäyttöön perustuvia sanakirjoja ja oppimateriaaleja.

Yhden mahdollisen – joskaan ei täydellisen – ratkaisun sopivien valmiiden oppimateriaalien puutteeseen tarjoavat korpuksat ja niihin liittyvät sähköiset työkalut. Esimerkiksi Korpin tyyppisellä verkkopohjaisella työkalulla tehty konkordanssihakua voi antaa sellaista tietoa sanan käyttöyhteyksistä, jota ei ole toistaiseksi saatavilla muualta. Perehtyminen korpuslingvististen työkalujen perustoimintoihin auttaisi sekä kielenopettajaa opetusmateriaaliensa laatimisessa että kielenoppijaa fraseologian tuntemuksensa kehittämässä. Kielenoppijoita kannattaisi siis totuttaa käyttämään ja kannustaa hyödyntämään niitä korpusaineistoja ja -työkaluja, jotka ovat vapaasti saatavilla.

Näyttää siis siltä, että yhtä yleispätevää tapaa sisällyttää fraseologiaa kielenopetukseen ei ole, vaan opetusta ja opetusmateriaaleja on räätälöitävä jokaisen kielenoppijan tai kielenoppijaryhmän tavoitteiden ja kielenkäyttöalueiden mukaan. Tämä tietää paljon työtä ja kokeilemista, eikä fraseologian tyhjentävä käsittely liene mahdollista missään kielenopetustilanteessa. Fraseologian tärkeyden painottamisen lisäksi olisi siis syytä antaa kielenoppijoille parempia työkaluja heidän fraseologisen tuntemuksensa itsenäiseen kehittämiseen. Tällainen tieto ja tuki olisivat tarpeen erityisesti niille, jotka opiskelevat kohdekieltään vieraana kielenä ilman runsasta ja säännöllistä kielisyötettä autenttisista lähteistä.

5. PÄÄTÄNTÖ

Tässä pro gradu -tutkielmassa olen kartoittanut ja luokitellut niitä fraseologisia epätyypillisyyksiä, jotka esiintyvät vielä edistyneen tason oppijansuomessa. Tutkimusaineistona olen käyttänyt Oulun yliopiston Kansainvälisen oppijansuomen korpuksen (ICLFI) C-tasoisiksi arvioituja raakatekstejä. Vertailuaineistona puolestaan toimivat Ylen suomenkielisen uutisarkiston osakorpukset vuosilta 2011–2018, joista teen konkordanssihakuja Kielipankin verkkopohjaisella Korp-työkalulla. Olen ensin valinnut tapaukseni sekä oman että yhden äidinkielen kielenkäyttäjän kieli-intuition perusteella, minkä jälkeen olen tarkistanut havaintojeni paikkansapitävyyttä vertailuaineiston, sanakirjan sekä kielipöytäkirjojen avulla. Luokitteluni perustuu Kuuluvaisen (2015) laatimaan fraseologisten virheiden luokitteluun, johon olen myös tehnyt useita lisäyksiä ja muokkauksia.

Siinä missä Kuuluvaisen (2015) kartoitus keskittyy oppijansuomen A- ja B-tasolla esiintyviin fraseologisiin epätyypillisyyksiin, oma kartoitukseni nostaa esiin myös C-tason fraseologisia epätyypillisyyksiä. Siksi katson, että tutkimukseni sekä täydentää Kuuluvaisen aiempaa tutkimusta että tarjoaa entistä laajemman luokittelumallin mahdollisten tulevien oppijansuomen fraseologisten tutkimusten käytettäväksi.

Tulosteni perusteella edistyneessä oppijansuomessa esiintyvät fraseologiset epätyypillisyydet sijoittuvat koko leksikaalisuus–morfosyntaktisuus-jatkumolle eli ne ilmenevät suppeimmillaan yksittäisissä sanoissa ja laajimmillaan lauseen laajuisissa kokonaisuuksissa. Yleisellä tasolla ne jakautuvat sananvalinnan, morfologian, lauseketason sekä lauseketason epätyypillisyyksien pääryhmiin.

Sananvalinnan epätyypillisyyksien pääryhmä on suurin, ja se jakautuu yksittäisten lähimerkityksisten sanojen, intensiteettipartikkelien, muiden kollokaatioiden ja leksikaalisten myötäesiintymien, semanttisen preferenssin ja semanttisen prosodian, uudismuodostusten, sanakirjan sudenkuoppien ja transfer-tapausten, virheellisten nomininvalintojen sekä virheellisten verbinvalintojen alaryhmiin. Niistä edelliset neljä alaryhmää sisältävät pääosin epätyypillisyyksiä, kun taas jälkimmäisen neljän alaryhmän tapaukset joko lähentelevät virheitä tai ovat suorastaan virheellisiä.

Alaryhmät puolestaan jakautuvat edelleen keskittymiin silloin, kun niissä on riittävästi tietyn yhdistävän piirteen jakavia tapauksia. Yksittäisten lähimerkityksisten sanojen alaryhmä jakautuu kontekstiin sopimattomien substantiivien, adjektiivien, kvanttorisanojen, yksittäisten verbien sekä modaaliverbien keskittymiin. Intensiteettipartikkelien alaryhmän keskittymät ovat puolestaan sanojen *aivan* ja *oikein* kollokaatio ja semanttinen preferenssi, sanan *sangen* kollokaatio, sanan *kovin* kolligaatio ja semanttinen preferenssi sekä sanojen *hyvin* ja *erittäin* kolligaatio. Muiden kollokaatioiden ja leksikaalisten myötäsiintymien alaryhmä jakautuu epätyypillisiin substantiivi-, adjektiivi-, adverb- sekä verbikollokaatteihin. Semanttisen preferenssin ja semanttisen prosodian alaryhmän keskittymät taas ovat epätyypillinen semanttinen preferenssi adjektiiveissa ja verbeissä sekä epätyypillinen semanttinen prosodia.

Uudismuodosteiden alaryhmä jakautuu yksi- ja useampisanaisten uudismuodosteiden keskittymiin. Sanakirjan sudenkuoppien ja transfer-tapausten alaryhmän keskittymät ovat puolestaan sanojen *liian* ja *kovin* sekoittaminen, fokuspartikkelien *jopa* ja *edes* sekoittaminen, sanojen *siis* ja *joten* sekoittaminen, yleis- ja erikoiskielen sanojen sekoittaminen, yksittäiset sopimattomat sanat sekä varsinaiset transfer-tapaukset. Virheellisten nomininvalintojen alaryhmä sisältää epätyypillisten tai virheellisten substantiivin- ja adjektiivinvalintojen keskittymät, kun taas virheellisten verbinvalintojen alaryhmä jakautuu transitiivi- ja intransitiiviverbien sekoittamisen sekä väärän johdoksen valinnan keskittymiin.

Morfologisten epätyypillisyyksien pääryhmä on pienin, ja se koostuu kolmesta keskittymästä, jotka ovat sanan epätyypillinen luku, sanan epätyypillinen sijapäätte sekä *sti*-adverbijohdosten ylimuodostus. Tapaukset tukevat Jantusen ja Brunnin (2012: 94–95) väitteitä siitä, että morfologinen priming ja fraseologia ovat yhteydessä toisiinsa ja että suomenoppijoiden morfologisen primingin hallinta poikkeaa jossain määrin natiivien morfologisen primingin hallinnasta.

Lauseketason epätyypillisyyksien pääryhmä muodostuu substantiivi-, adjektiivi- ja partiisiippi-, adpositio- sekä infinitiivilausekkeiden alaryhmistä. Substantiivilausekkeiden alaryhmä sisältää sananvalinnan ja sananmuodon muutoksia, lausekkeen rakenteen muutoksia sekä isompia syntaktisia muutoksia vaativien epätyypillisyyksien keskittymät. Adjektiivi- ja partiisiippilausekkeiden alaryhmä jakautuu puolestaan kahteen keskittymään nimensä mukaan. Adpositiolausekkeiden alaryhmän keskittymät ovat adposition käyttö

sijapäänteen tilalla ja vastaavasti sijapäänteen käyttö adposition tilalla. Infinitiivilausekkeiden alaryhmä jakautuu taas epätyypillisten A-, E- ja MA-infinitiivilausekkeiden keskittymiin.

Lausetason epätyypillisyyksien pääryhmään kuuluvat verbin valenssin, verbin ja sen täydennysympäristön, lausetyyppien, *tulla joksikin* -rakenteen, vierasperäisten lauserakenteiden sekä kiteytyneiden ilmausten alaryhmät. Näistä alaryhmistä lausetyyppien, *tulla joksikin* -rakenteen ja kiteytyneiden ilmausten alaryhmät jakautuvat edelleen keskittymiin. Lausetyyppien alaryhmä koostuu omistuslauseen ja nesessiivirakenteen epätyypillisen käytön keskittymistä. *Tulla joksikin* -rakenteen alaryhmän keskittymät taas ovat rakenteen käyttö tuloslauseen sijasta, rakenteen käyttö refleksiivisen verbijohdoksen tilalla sekä *tulla*-verbin käyttö translatiivimuotoisen predikatiiviadverbiaalin yhteydessä. Kiteytyneiden ilmausten alaryhmän keskittymät ovat puolestaan rakenteen epäonnistunut muodostaminen sekä onnistuneesti muodostetun rakenteen epätyypillinen käyttö.

Alaryhmien ja keskittymien paljoudesta näkyy, että edistyneen oppijansuomen fraseologisten epätyypillisyyksien joukossa on yksittäistapausten lisäksi myös useita yleistettävissä olevia piirteitä. Tutkimustulokseni antavat kattavan, joskaan eivät tyhjentävän yleiskuvan näistä yleisimmistä tapauksista. Niiden suuri määrä ja monipuolisuus osoittavat, että suomenoppijoilta on odotettavissa runsaasti fraseologisia epätyypillisyyksiä edistyneelläkin tasolla. Lisäksi nämä epätyypillisyydet voivat ilmetä yllättävänkin yleisten ja taajaan käytettyjen sanojen ja rakenteiden yhteydessä; ne eivät siis koske pelkästään harvinaisempia, ”edistyneemmän tason” sanoja ja rakenteita.

Vaikka olenkin tutkimuksen alkuvaiheessa toiminut pääosin kieli-intuitioni varassa poimiessani epätyypillisyyksiä, pidän johtopäätöksiäni luotettavina sillä perusteella, että niiden tukena on runsas, todelliseen kielenkäyttöön perustuva korpusaineisto, jolle olen myös antanut ensisijaisen painoarvon. Lisäksi alustavat havaintoni eivät ole syntyneet pelkästä introspektiosta, vaan nekin pohjautuvat autenttiseen kielilähteeseen, ICLFIn oppijansuomen korpusaineistoon. Koska olen luetellut kaikki konkordanssihaussa käyttämäni kyselylausekkeet tämän tutkielman liitteessä, tulosteni paikkansapitävyyttä on myös helppo testata Korp-työkalun avulla. Tutkimustulosteni yhteensopivuutta muiden samanaikaisen tutkimusten kanssa puoltaa se, että valitsemani tapaukset sijoittuvat hyvin Kuuluvaisen (2015) kehittämään luokitusmalliin.

Tulokseni tarjonnevat siis tietoa siitä, mihin osa-alueisiin olisi syytä kiinnittää huomiota ja panostaa erityisesti suomi vieraana kielenä -opetuksessa, jossa suomenoppijalla saattaa olla rajatusti kosketuksia kohdekieleensä. Lisäksi ne puoltavat useiden tutkijoiden näkemystä siitä, että sanastoa on opetettava fraseologisina kokonaisuuksina jo heti opetuksen alussa. Mikäli tällaisiin ongelmakohtiin puututaan riittävän ajoissa, voidaan ehkä päästä lähemmäs niitä kielellisten kompetenssien kriteereitä, joita Eurooppalainen viitekehys (2003) on asettanut C-tason kielenoppijoille.

ICLFI on kirjoitushetkellä ainoa seitsemästä oppijansuomen korpuksesta, joka edustaa suomi vieraana kielenä -varianttia (Jantunen & Pirkola 2015: 91). Jatkossa olisi mielenkiintoista vertailla löytämiäni epätyypillisyyksiä niihin epätyypillisyyksiin, joita esiintyy suomi toisena kielenä -variantissa. Kuuluvaiselta (2015) lainaamani ja päivittämäni fraseologisten epätyypillisyyksien luokittelumalli soveltunee hyvin tällaiseen tutkimukseen, kuten myös saman aiheen laajempaan tutkimukseen, jossa kartoitettaisiin ja tarkasteltaisiin fraseologisia epätyypillisyyksiä useammalta taitotasolta. Toisenlainen tutkimus, jonka tämä pääpiirteittäinen kartoitus voisi poikia, on tietyn epätyypillisyyden alaryhmän tai keskittymän syvällisempi ja tilastopainotteisempi tarkastelu sekä vertailu natiivisuomeen. Erityisesti oppijansuomen morfologiseen tarkasteluun olisi syytä kiinnittää ennististä enemmän huomiota, sillä alustavat löydökseni ovat vain pintaraapaisu toistaiseksi vähän tutkitusta oppijansuomen osa-alueesta.

Kaiken kaikkiaan voidaan todeta, että siinä missä puutteellinen fraseologinen hallinta saattaa vaikeuttaa ymmärrettävyyttä alkeis- tai jopa keskitasolla, se näyttää olevan lähinnä kompastuskivi niille edistyneille suomenoppijoille, jotka syystä tai toisesta haluavat sulautua kielellisesti natiivien joukkoon. Toisaalta natiivilähtöisen kielenkäyttömallin sopivuutta kielenoppijoille on aiemmin kyseenalaistettu, eivätkä kaikki kielenoppijat edes tavoittele natiivinkaltaista kielitaitotasoa (Cook 1999). Tämä pätee myös suomenoppijoihin, mutta tällaiselle tasolle pyrkivät – kuten tämän tutkielman kirjoittaja – eivät pääse sinne ilman fraseologian vankkaa tuntemusta ja hallintaa.

LÄHTEET JA LYHENTEET

Lyhenteet

ADJ = adjektiivi

ADV = adverbi

A-INF = A-infinitiivi

AKT = aktiivi

ALL = allatiivi

ILL = illatiivi

IMPF = imperfekti

INE = inessiivi

MON = monikko

OBJ = objekti

SUBJ = subjekti

SUBST = substantiivi

VER = verbi

YKS = yksikkö

Tutkimusaineistot

ICLFI = International Corpus of Learner Finnish eli Kansainvälisen oppijansuomen korpus. Saatavilla <http://urn.fi/urn:nbn:fi:lb-20140730163>. Oppijansuomen sähköinen tutkimusaineisto. Koostettu hankkeessa Korpustutkimus oppijankielen kielikohtaisista ja universaaleista ominaisuuksista Oulun yliopiston suomi toisena ja vieraana kielenä -oppi-aineessa.

<https://www oulu.fi/suomitoisenakielenä/node/16078> (20.4.2021).

Ylen suomenkielinen uutisarkisto 2011–2018. Korp [tekstikorpus]. Kielipankki. Saatavilla <http://urn.fi/urn:nbn:fi:lb-2019121003>. Sisältää myös Ylen suomenkielisen uutisarkiston selkouutiset samoilta vuosilta.

<https://www.kielipankki.fi/aineistot/ylenews/> (20.4.2021).

Kirjallisuus

- AKGÜL, ANNA 2013: *Frekventtien adjektiivien kollokaatiot oppijansuomessa ja natiivisuomessa*. Pro gradu -tutkielma. Oulun yliopisto, suomen kieli.
- ALDERSON, J. CHARLES 2007: The CEFR and the need for more research. – *The Modern Language Journal* 91 (4) s. 659–663.
- ARPPE, ANTTI 2008: *Univariate, bivariate and multivariate methods in corpus-based lexicography. A study of synonymy*. Väitöskirja. Publications of the department of general linguistics no. 44. Helsingin yliopisto.
- BARLOW, MICHAEL 2005: Computer-based analyses of learner language. – Rod Ellis & Gary Barkhuizen (toim.), *Analysing learner language* s. 335–357. Oxford: Oxford University Press.
- BEDNAREK, MONIKA 2008: Semantic preference and semantic prosody re-examined. – *Corpus Linguistics and Linguistic Theory* 4 (2) s. 119–139.
- BIBER, DOUGLAS – CONRAD, SUSAN 1999: Lexical bundles in conversation and academic prose. – Hilde Hasselgård & Signe Oksefjell (toim.), *Out of corpora. Studies in honour of Stig Johansson* s. 181–190. Amsterdam: Rodopi.
- BIBER, DOUGLAS – CONRAD, SUSAN – REPPEN, RANDI 1998: *Corpus linguistics. Investigating language structure and use*. Cambridge Approaches to Linguistics. Cambridge: Cambridge University Press.
- BORIN, LARS – FORSBERG, MARKUS – ROXENDAL, JOHAN 2012: Korp. The corpus infrastructure of Språkbanken. *Proceedings of LREC 2012* s. 474–478. Istanbul: ELRA.
- BRUNNI, SISCO – JANTUNEN, JARMO H. 2015: Learner language. – Martha Bigelow & Johanna Ennser-Kananen (toim.), *The Routledge handbook of educational linguistics* s. 383–394. New York: Routledge.
- BRUNNI, SISCO – JANTUNEN, JARMO H. – SKANTSI, VALTTERI 2020: Korpusavusteinen virheanalyysi tarkkuuden kehityksestä EVK:n taitotasoilla A2–B2. – *Puhe ja kieli* 39 (3) s. 275–304.
- CHAFE, WALLACE 1992: The importance of corpus linguistics to understanding the nature of language. – Jan Svartvik (toim.), *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991* s. 79–97. Berlin: De Gruyter Mouton.
- CHENG, YING XI 2014: *Fraseologia suomenoppijan kompastuskivenä*. Kandidaatintutkielma. Oulun yliopisto, suomen kieli.

- COOK, VIVIAN 1999: Going beyond the native speaker in language teaching. – *TESOL Quarterly* 33 (2) s. 185–209.
- CORDER, S. PIT 1967: The significance of learners' errors. – *International Review of Applied Linguistics* 5 (4) s. 161–170.
- 1973: *Introducing applied linguistics*. Baltimore: Penguin.
- 1974: Error analysis. – John P. B. Allen & S. Pit Corder (toim.), *Techniques in applied linguistics* s. 122–154. The Edinburgh Course in Applied Linguistics Volume 3. Oxford: Oxford University Press.
- 1981: *Error analysis and interlanguage*. Oxford: Oxford University Press.
- CRUSE, D. ALAN 1986: *Lexical semantics*. Cambridge: Cambridge University Press.
- 2000: *Meaning in language. An introduction to semantics and pragmatics*. Oxford: Oxford University Press.
- DAGNEAUX, ESTELLE – DENNESS, SHARON – GRANGER, SYLVIANE 1998: Computer-aided error analysis. – *System* 26 (2) s. 163–174.
- EKI ühendsõnastik 2021. Eesti Keele Instituut, Sõnaveeb. <https://sonaveeb.ee/> (20.4.2021).
- EKSS = *Eesti keele seletav sõnaraamat* 2009. Eesti Keele Instituut. <http://www.eki.ee/dict/ekss/> (20.4.2021).
- ELLIS, NICK C. 2008: Phraseology. The periphery and the heart of language. – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in foreign language learning and teaching* s. 1–13. Amsterdam: John Benjamins.
- ELLIS, ROD 2008: *The study of second language acquisition*. 2. painos. Oxford: Oxford University Press.
- ELLIS, ROD – BARKHUIZEN, GARY 2005: *Analysing learner language*. Oxford: Oxford University Press.
- ERELT, MATI – ERELT, TIJU – ROSS, KRISTIINA 2007: *Eesti keele käsiraamat*. 3. painos. Tallinn: Eesti Keele Sihtasutus.
- EVK = *Eurooppalainen viitekehys. Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys*. Porvoo: WSOY 2003.
- FIRTH, JOHN R. 1957: *Papers in linguistics 1934–1951*. London: Oxford University Press.
- 1968 [1957]: A synopsis of linguistic theory 1930–1955. – F. R. Palmer (toim.), *Selected papers of J. R. Firth 1952–59* s. 168–205. Bloomington: Indiana University Press.

- FRANCIS, GILL 1993: A corpus-driven approach to grammar. Principles, methods and examples. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology. In honour of John Sinclair* s. 137–156. Amsterdam: John Benjamins.
- GRANGER, SYLVIANE 1998a: Prefabricated patterns in advanced EFL writing. Collocations and formulae. – Anthony P. Cowie (toim.), *Phraseology: Theory, analysis and applications* s. 145–160. Oxford: Oxford University Press.
- 1998b: The computer learner corpus. A versatile new source of data for SLA research. – Sylviane Granger (toim.), *Learner English on computer* s. 3–18. London: Longman.
- 2002: A bird's-eye view of learner corpus research. – Sylviane Granger, Joseph Hung & Stephanie Petch-Tyson (toim.), *Computer learner corpora, second language acquisition and foreign language teaching* s. 3–33. Amsterdam: John Benjamins.
- 2008: Learner corpora. – Anke Lüdeling & Merja Kytö (toim.), *Corpus Linguistics. An International Handbook. Volume 1* s. 259–275. Handbooks of Linguistics and Communication Science, Band 29.1. Berlin: De Gruyter Mouton.
- GRANGER, SYLVIANE – MEUNIER, FANNY 2008: Introduction. The many faces of phraseology. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. xix–xxviii. Amsterdam: John Benjamins.
- GRANGER, SYLVIANE – PAQUOT, MAGALI 2008: Disentangling the phraseological web. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. 27–49. Amsterdam: John Benjamins.
- GRIES, STEFAN TH. 2002: Evidence in linguistics. Three approaches to genitives in English. – Brend, Ruth M., William J. Sullivan and Arle R. Lommel (toim.), *LACUS Forum XXVIII. What constitutes evidence in linguistics?* s. 17–31. Fullerton: LACUS.
- 2008: Phraseology and linguistic theory. A brief survey. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology. An interdisciplinary perspective* s. 3–25. Amsterdam: John Benjamins.
- 2009: What is corpus linguistics? – *Language and Linguistics Compass* 3 s. 1–17.
- HALLIDAY, M. A. K. 1966: Lexis as a Linguistic Level. – C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (toim.), *In memory of J. R. Firth* s. 148–162. London: Longman.

- HARDIE, ANDREW – MCENERY, TONY 2010: On two traditions in corpus linguistics, and what they have in common. – *International Journal of Corpus Linguistics* 15 (3) s. 384–394.
- HEIKKINEN, VESA 1999: *Ideologinen merkitys kriittisen tekstintutkimuksen teoriassa ja käytännössä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HEIKKINEN, VESA – LOUNELA, MIKKO – VOUTILAINEN, EERO 2012: Automaattinen analysaattori tekstilajitutkimuksessa. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi. Tekstilajitutkimuksen käsikirja* s. 372–391. Kotimaisten kielten keskuksen julkaisuja 169. Helsinki: Gaudeamus.
- HOEY, MICHAEL 1991: *Patterns of lexis in text*. Oxford: Oxford University Press.
- 2005: *Lexical priming. A new theory of words and language*. London: Longman.
- HUNSTON, SUSAN 2006: Corpus linguistics. – Keith Brown (toim.), *Encyclopedia of language and linguistics* s. 234–248. Amsterdam: Elsevier.
- HÄKKINEN, KAISA 1994: *Kielitieteen perusteet*. Tietolipas 133. Helsinki: Suomalaisen Kirjallisuuden Seura.
- IKÄVALKO, RIIKKA 2019: *Hirveän ja kauhean käyttö ja merkitykset oppijansuomessa*. Pro gradu -tutkielma. Turun yliopiston kieli- ja käännöstieteiden laitos.
- ISOHÄTÄLÄ, JAANA 2014: *Kielenoppimisympäristön vaikutus oppijansuomen sanastoon*. Pro gradu -tutkielma. Suomen kieli, Oulun yliopisto.
- IVASKA, ILMARI 2015: *Edistyneen oppijansuomen konstruktiopiirteitä korpusvetoisesti. Avainrakenneanalyysi*. Väitöskirja. Turun yliopiston kieli- ja käännöstieteiden laitos.
- JAMES, CARL 1998: *Errors in language learning and use. Exploring error analysis*. London: Longman.
- JANTUNEN, JARMO H. 2001: ”Tärkeä seikka” ja ”keskeinen kysymys”. Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista? – *Virittäjä* 105 (2) s. 170–192.
- 2004: *Synonymia ja käännössuomi. Korpusnäkökulma samanmerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaaliin erityispiirteisiin*. Väitöskirja. Joensuun yliopiston humanistisia julkaisuja 35. Joensuun yliopisto.
- 2008: Haasteita oppijankielen korpusanalyysille. Oppijankielen universaalit. – Pille Eslon (toim.), *Õppijakeele analüüs: võimalused, probleemid, vajadused* s. 1–26. Tallinna ülikooli eesti filoloogia osakonna toimetised 10. Tallinn: Tallinna Ülikooli Kirjastus.

- 2009a: Ei pelkästään mielikuvituksen puutteen vuoksi. Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. – *Virittäjä* 113 (1) s. 101–113.
- 2009b: Minulla on aivan paljon rahaa. Fraseologiset yksiköt suomen kielen opetuksessa. – *Virittäjä* 113 (3) s. 356–381.
- 2011: Kansainvälinen oppijansuomen korpus (ICLFI). Typologia, taustamuuttujat ja annotointi. – *Lähivõrdlusi. Lähivertailuja* 21 s. 86–105.
- 2012a: Akateemiset ja populaaritekstit korpusvertailussa. Sanalistat, avainsanat ja fraseologiset yksiköt. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi. Tekstilajitutkimuksen käytäntöä* s. 186–213. Kotimaisten kielten keskuksen verkkojulkaisuja 29. Helsinki: Kotimaisten kielten keskus.
- 2012b: Korpusvetoinen tekstilajianalyysi. Sanalistat ja genreavainsanat. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi. Tekstilajitutkimuksen käsikirja* s. 360–371. Kotimaisten kielten keskuksen julkaisuja 169. Helsinki: Gaudeamus.
- JANTUNEN, JARMO H. – BRUNNI, SSKO 2012: Morfologinen priming ja fraseologia vieraan kielen oppimisessa. Korpustutkimus oppijansuomesta. – *Lähivõrdlusi. Lähivertailuja* 22 s. 71–100.
- JANTUNEN, JARMO H. – BRUNNI, SSKO – LEHTO, LIISA-MARIA – AIRAKSINEN, VALTTERI 2014: Oppijankieliaineistojen annotointi. Esimerkkinä ICLFI:n annotoinnin prosessit, ongelmat ja ratkaisut. – Maarit Mutta, Pekka Lintunen, Ilmari Ivaska & Pauliina Peltonen (toim.), *AFinLA-e. Soveltavan kielitieteen tutkimuksia* 7 s. 60–80. Jyväskylä: AFinLa.
- JANTUNEN, JARMO – PILTONEN, SAANA 2009: Oppijansuomen ja -viron sähköiset tutkimusaineistot. – *Virittäjä* 113 (3), s. 449–458.
- JANTUNEN, JARMO H. – PIKOLA, SILJA 2015: Oppijansuomen sähköiset tutkimusaineistot. Nykytilanne. – *Virittäjä* 119 (1), s. 88–103.
- KALLIORANTA, OTTO 2009: *Paljon-adverbin kollokointi oppijansuomessa. Korpusvetoinen tutkimus*. Pro gradu -tutkielma. Oulun yliopisto, suomen kieli.
- KANGAS, AKSELI 2018: *Adverbien täysin ja kokonaan kollokaatit ja semanttinen prosodia*. Pro gradu -tutkielma. Jyväskylän yliopiston kieli- ja viestintätieteiden laitos.
- KANGASNIEMI, HEIKKI 1997: *Sana, merkitys, maailma. Katsaus leksikaalisen semantiikan perusteisiin*. Helsinki: Finn Lectura.

- KARLSSON, FRED 2008: *Yleinen kielitiede*. Uudistetun laitoksen 5. painos. Helsinki: Gaudamus.
- KENNEDY, GRAEME 2008: Phraseology and language pedagogy. Semantic preference associated with English verbs in the British National Corpus. – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in foreign language learning and teaching* s. 21–41. Amsterdam: John Benjamins.
- KIELIPANKKI 2020a: Etusivu. – <https://www.kielipankki.fi/> (20.4.2021).
- 2020b: Ylen uutisaineistot. – <https://www.kielipankki.fi/aineistot/ylenews/> (20.4.2021).
- KS = *Kielitoimiston sanakirja*. 2020. Helsinki: Kotimaisten kielten keskuksen verkkojulkaisuja 35. URN:NBN:fi:kotus-201433. <URL <https://www.kielitoimistonsanakirja.fi>>. Päivitettävä julkaisu. Päivitetty 11.11.2020.
- KUULUVAINEN, HELENA 2015: *Fraseologiset virheet kansainvälisessä oppijansuomen korpuksessa*. Pro gradu -tutkielma. Oulun yliopisto, suomen kieli.
- LARSEN-FREEMAN, DIANE – LONG, MICHAEL H. 1991: *An introduction to second language acquisition research*. London: Longman.
- LATOMAA, SIRKKU 1993: Mitä hyötyä on oppijoiden kielitaustan tuntemisesta? – Eija Aalto & Minna Suni (toim.), *Kohdekielenä suomi. Näkökulmia opetukseen* s. 9–31. Korkeakoulujen kielikeskusten selosteita 1. Jyväskylän yliopisto, Korkeakoulujen kielikeskus.
- LATOMAA, SIRKKU – TUOMELA, VELI 1993: Suomi toisena vai vieraana kielenä? – *Viritäjä* 97 (2) s. 238–245.
- LEECH, GEOFFREY 1981: *Semantics. The study of meaning*. 2. painos. Harmondsworth: Penguin Books.
- 1993: Corpus annotation schemes. *Literary and Linguistic Computing* 8 (4) s. 275–281.
- LINDROOS-ČERMÁKOVÁ, HILKKA 1993: Fraseologiaa suomea vieraana kielenä opiskelevillekin. – Jyrki Kalliokoski & Kirsti Siitonen (toim.), *Suomeksi maailmalla. Kirjoituksia suomen kielen ja kulttuurin opettamisesta* s. 87–96. Castrenianumin toimitteita 44. Helsinki: Yliopistopaino.
- LITTLE, DAVID 2007: The Common European Framework of Reference for Language. Perspectives on the making of supranational language education policy. – *The Modern Language Journal* 91 (4) s. 645–655.

- LOUW, BILL 1993: Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology. In honour of John Sinclair* s. 157–176. Amsterdam: John Benjamins.
- LÜDELING, ANKE – KYTÖ, MERJA 2008: Introduction. – Anke Lüdeling & Merja Kytö (toim.), *Corpus linguistics. An international handbook. Volume 1*. Handbooks of Linguistics and Communication Science, Band 29.1. Berlin: De Gruyter Mouton.
- LYONS, JOHN 1968: *Introduction to theoretical linguistics*. London: Cambridge University Press.
- MARTIN, MAISA 1995: *The map and the rope. Finnish nominal inflection as a learning target*. Väitöskirja. Studia Philologica Jyväskyläensia 38. Jyväskylän yliopisto.
- MARTIN, MAISA – ALANEN, RIIKKA 2011: Oppijankieli. Kehitystä ja virheitä. – Paula Kallaja, Riikka Alanen & Hannele Dufva (toim.), *Kieltä tutkimassa. Tutkielman laatijan opas* s. 33–48. Helsinki: Finn Lectura.
- MAURANEN, ANNA 2004: Spoken corpus for an ordinary learner. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 89–105. Studies in Corpus Linguistics 12. Amsterdam: John Benjamins.
- 2011: Learners and users. Who do we want corpus data from? – Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin & Magali Paquit (toim.), *A taste for corpora. In honour of Sylviane Granger*. Studies in Corpus Linguistics 45. Amsterdam: John Benjamins.
- MEUNIER, FANNY – GRANGER, SYLVIANE 2008: Phraseology in language learning and teaching. Where to from here? – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in foreign language learning and teaching* s. 247–252. Amsterdam: John Benjamins.
- MOT Kotus suomi–viro-suursanakirja*. Kielikone Oy.
- MOT Sanakirjat*. Kielikone Oy. <https://www.sanakirja.fi/> (20.4.2021).
- MOT Venäjä*. Kielikone Oy.
- NENONEN, MARJA 2002: *Idiomit ja leksikko. Lausekeidiomien syntaktisia, semanttisia ja morfologisia piirteitä suomen kielessä*. Väitöskirja. Joensuun yliopiston humanistisia julkaisuja 29. Joensuun yliopisto.
- NESSERHAUF, NADJA 2003: The use of collocations by advanced learners of English and some implications for teaching. – *Applied Linguistics* 24 (2) s. 223–242.

- 2004a: What are collocations? – David Allerton, Nadja Nesselhauf, & Paul Skandera (toim.), *Phraseological units. Basic concepts and their application* s. 1–21. Basel: Schwabe.
- 2004b: Learner corpora and their potential for language teaching. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 125–152. Studies in Corpus Linguistics 12. Amsterdam: John Benjamins.
- 2005: *Collocations in a learner corpus*. Studies in Corpus Linguistics 14. Amsterdam: John Benjamins.
- NISSILÄ, LEENA 2011: *Viron kielen vaikutus suomen kielen verbien ja niiden rektioiden oppimiseen*. Väitöskirja. Acta Universitatis Ouluensis B Humaniora 99. Oulun yliopisto, suomen kieli.
- ODLIN, TERENCE 1989: *Language transfer. Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- OJUTKANGAS, KRISTA 2008: Kielentutkijan mielikuvituksen puute eli mihin todelliseen kielenkäyttöön perustuvaa aineistoa tarvitaan. – *Virittäjä* 112 (2) s. 276–284.
- OPETUSHALLITUS 2003: *Lukion opetussuunnitelman perusteet 2003. Nuorille tarkoitetun lukiokoulutuksen opetussuunnitelman perusteet*. https://www.oph.fi/sites/default/files/documents/47345_lukion_opetussuunnitelman_perusteet_2003.pdf (16.5.2021).
- PARTINGTON, ALAN 1998: *Patterns and meanings. Using corpora for English language research and teaching*. Studies in Corpus Linguistics 2. Amsterdam: John Benjamins.
- 2004: Utterly content in each other's company. Semantic prosody and semantic preference. *International Journal of Corpus Linguistics* 9 (1) s. 131–156.
- PIRKOLA, SILJA 2016: *Synonyymien EHKÄ ja MAHDOLLISESTI kollokaatit ja semanttiset preferenssit*. Pro gradu -tutkielma. Jyväskylän yliopiston kieli- ja viestintätieteiden laitos.
- RICHARDS, JACK C. 1974: A non-contrastive approach to error analysis. – Jack C. Richards (toim.), *Error analysis. Perspectives on second language acquisition* s. 172–188. London: Longman.
- SAJAVAARA, KARI 1999: Kontrastiivinen kielentutkimus ja virheanalyysi. – Kari Sajaavaara & Anja Piirainen-Marsh (toim.), *Kielenoppimisen kysymyksiä* s. 103–128. Soveltavan kielentutkimuksen teoriaa ja käytäntöä 7. Jyväskylän yliopiston soveltavan kielentutkimuksen keskus.

- Sanakirja.org*. Sanakirja.org Solutions Oy. <https://www.sanakirja.org/> (20.4.2021).
- SCHACHTER, JAQUELYN – CELCE-MURCIA, MARIANNE 1977: Some reservations concerning error analysis. – *TESO Quarterly* 11 s. 441–451.
- SCHOLFIELD, PHIL 1995: *Quantifying Language. A researcher's and teacher's guide to gathering language data and reducing it to figures*. Clevedon: Multilingual Matters.
- SCHÜTZE, CARSON T. 2016: *The empirical base of linguistics. Grammaticality judgements and linguistic methodology*. Classics in Linguistics 2. Berlin: Language Science Press.
- SELINKER, LARRY 1969: Language transfer. *General Linguistics* 9 (2) s. 67–92.
- 1972: Interlanguage. *International Review of Applied Linguistics* 10 (3) s. 209–231.
- SEPPÄLÄ, TANJA 2012: *Oppijansuomen kolligaatit ketjuuntuviissa verbirakenteissa*. Pro gradu -tutkielma. Oulun yliopisto, suomen kieli.
- SINCLAIR, JOHN 1991: *Corpus, concordance, collocation. Describing English language*. Oxford: Oxford University Press.
- 1996: The search for units of meaning. – *Textus* 9 (1) s. 75–106.
- 1998: The lexical item. – Weigand, Edda (toim.), *Contrastive lexical semantics* s. 1–24. Current Issues in Linguistic Theory 171. Amsterdam: John Benjamins.
- 2004a: Introduction. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 1–10. Studies in Corpus Linguistics 12. Amsterdam: John Benjamins.
- 2004b: New evidence, new priorities, new attitudes. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 271–299. Studies in Corpus Linguistics 12. Amsterdam: John Benjamins.
- STUBBS, MICHAEL 1995: Collocations and semantic profiles. On the cause of the trouble with quantitative studies. – *Functions of Language* 2 (1) s. 23–55. Amsterdam: John Benjamins.
- 1996: *Text and corpus analysis. Computer-assisted studies of language and culture*. Language in Society 23. Oxford: Blackwell.
- 2001a: *Words and phrases. Corpus studies of lexical semantics*. Oxford: Blackwell.
- 2001b: On inference theories and code theories. Corpus evidence for semantic schemas. – *Text* 21 (3) s. 437–465.
- 2002: On text and corpus analysis. A reply to Borsley and Ingham. – *Lingua* 112 (1) s. 7–11.

- SUORANIEMI, HENRIKKA 2011: *Vuoksi ja takia. Korpuspohjainen tutkimus lähisynonyymeistä*. Pro gradu -tutkielma. Oulun yliopisto, suomen kieli.
- TARVAINEN, JENNY 2018: *SAADA-verbin fraseologiaa. Vertaileva korpustutkimus oppijanja natiivisuomesta*. Pro gradu -tutkielma. Jyväskylän yliopiston kieli- ja viestintätieteiden laitos.
- TOGNINI-BONELLI, ELENA 2001: *Corpus linguistics at work*. Studies in Corpus Linguistics 6. Amsterdam: John Benjamins.
- TSUI, AMY B. M. 2004: What teachers have always wanted to know – and how corpora can help. – John Sinclair (toim.), *How to use corpora in language teaching* s. 39–61. Studies in Corpus Linguistics 12. Amsterdam: John Benjamins.
- TUMMERS, JOSE – HEYLEN, KRIS – GEERAERTS, DIRK 2005: Usage-based approaches in cognitive linguistics. A technical state of the art. *Corpus linguistics and linguistic theory* 1 (2) s. 225–261.
- VARRIO, HANNA 2014: *Sanoa-verbi fraseologisena yksikkönä oppijansuomessa ja natiivikielessä. Verbin ydin ja täydennysympäristö*. Pro gradu -tutkielma. Oulun yliopisto, suomen kieli.
- VISK = HAKULINEN, AULI – VILKUNA, MARIA – KORHONEN, RIITTA – KOIVISTO, VESA – HEINONEN, TARJA RIITTA – ALHO, IRJA 2004: *Iso suomen kielioppi*. Verkko-versio. Helsinki: Suomalaisen Kirjallisuuden Seura. <http://scripta.kotus.fi/visk> (20.4.2021).
- WASOW, THOMAS – ARNOLD, JENNIFER 2005: Intuitions in linguistic argumentation. – *Lingua* 115 s. 1481–1496.
- YLEISRADIO 2021a: Ylen suomenkielinen uutisarkisto 2011–2018, Korp [tekstikorpus]. Kielipankki. Saatavilla <http://urn.fi/urn:nbn:fi:lb-2019121003> (20.4.2021).
- 2021b. Ylen suomenkielisen uutisarkiston selkouutiset 2011–2018, Korp [tekstikorpus]. Kielipankki. Saatavilla <http://urn.fi/urn:nbn:fi:lb-2019121203> (20.4.2021).

LIITE. CQP (*Corpus Query Protocol*) -kyselylausekkeet aakkosjärjestyksessä.

Luku	Hakusana tai -yhdistelmä	CQP-kyselylauseke
4.2.1.	AATE	[lemma = "aate"]
4.5.1.	AIHEUTTAA + [0–1 sana] + ETTÄ	[lemma = "aiheuttaa"] [] {0,1} [lemma = "että"]
4.5.1.	AIHEUTTAA + SE + [0–1 sana] + ETTÄ	[lemma = "aiheuttaa"] [lemma = "se"] [] {0,1} [lemma = "että"]
4.2.2.	AIVAN + MUU	[lemma = "aivan"] [lemma = "muu"]
4.2.6.	AJANKOHTAINEN	[lemma = "ajankohtainen"]
4.2.6.	AKTUAALINEN	[lemma = "aktuaalinen"]
4.5.2.	ANSAITA	[lemma = "ansaita"]
4.5.2.	ANSAITA & -MINEN	[msd = ".*minen.*"] [lemma = "ansaita"]
4.2.2.	AIVAN + ADJ	[lemma = "aivan"] [pos = "A"]
4.2.2.	AIVAN + MIELENKIINTOINEN	[lemma = "aivan"] [lemma = "mielenkiintoinen"]
4.2.2.	AIVAN + TÄRKEÄ	[lemma = "aivan"] [lemma = "tärkeä"]
4.2.1.	EI + MALTTAA + ODOTTAA	[lemma = "ei"] [lemma = "maltaa"] [lemma = "odottaa"]
4.2.2.	EI + OLLA + KOVIN + VAHVA	[lemma = "ei"] [lemma = "olla"] [lemma = "kovin"] [lemma = "vahva"]
4.2.2.	EI + OLLA + OIKEIN + VAHVA	[lemma = "ei"] [lemma = "olla"] [lemma = "oikein"] [lemma = "vahva"]
4.2.2.	EI + VER + KOVIN + USEIN	[lemma = "ei"] [pos = "V"] [lemma = "kovin"] [lemma = "usein"]
4.2.2.	EI + VER + OIKEIN + USEIN	[lemma = "ei"] [pos = "V"] [lemma = "oikein"] [lemma = "usein"]
4.2.1.	EI + VOIDA + ODOTTAA	[lemma = "ei"] [lemma = "voida"] [lemma = "odottaa"]

Luku	Hakusana tai -yhdistelmä	CQP-kyselylauseke
4.2.4.	ELÄMYS	[lemma = "elämys"]
4.2.2.	ERITTÄIN + TÄRKEÄ	[lemma = "erittäin"] [lemma = "tärkeä"]
4.2.2.	ERITTÄIN + TÄRKEÄ	[lemma = "erittäin"] [lemma = "tärkeä"]
4.2.6.	FAKTORI	[lemma = "faktori"]
4.2.4.	HANKALA + ASIAKAS	[lemma = "hankala"] [lemma = "asiakas"]
4.2.4.	HANKALA + IHMINEN	[lemma = "hankala"] [lemma = "ihminen"]
4.2.4.	HANKALA + TYÖNTEKIJÄ	[lemma = "hankala"] [lemma = "työntekijä"]
4.3.	<i>hirveän</i> + ADJ	[word = "hirveän"] [pos = "A"]
4.3.	HIRVEÄSTI + ADJ	[lemma = "hirveästi"] [pos = "A"]
4.2.2.	HYVIN + MIELENKIINTOINEN	[lemma = "hyvin"] [lemma = "mielenkiintoinen"]
4.2.2.	HYVIN + TÄRKEÄ	[lemma = "hyvin"] [lemma = "tärkeä"]
4.2.4.	HÄVITTÄÄ	[lemma = "hävittää"]
4.2.5.	ITSEMURHA + TEKIJÄ	[lemma = "itsemurha"] [lemma = "tekijä"]
4.2.5.	ITSEMURHAAJA	[lemma = "itsemurhaaja"]
4.2.4.	JOUTUA	[lemma = "joutua"]
4.2.3.	JÄRJESTÄÄ & ASIA (OBJ)	[lemma = "järjestää"] [[lemma = "asia" & deprel = "dobj"]]
4.2.3.	JÄRJESTÄÄ & SEIKKA (OBJ)	[lemma = "järjestää"] [[lemma = "seikka" & deprel = "dobj"]]
4.2.3.	JÄÄDÄ & ELÄKE (ALL)	[lemma = "jäädä"] [[msd = ".*All.*" & lemma = "eläke"]]
4.3.	<i>kaikesta</i> + <i>huolimatta</i>	[word = "kaikesta"] [word = "huolimatta"]
4.3.	<i>kaikista</i> + <i>huolimatta</i>	[word = "kaikista"] [word = "huolimatta"]
4.2.4.	KALPEA + SUBST	[lemma = "kalpea"] [pos = "N"]
4.5.2.	KANNATTAA + LISÄTÄ	[lemma = "kannattaa"] [lemma = "lisätä"]
4.2.4.	KARKOTTAA	[lemma = "karkottaa"]

Luku	Hakusana tai -yhdistelmä	CQP-kyselylauseke
4.2.3.	KERTOJA & MIELIPIDE (OBJ)	[lemma = "kertoa"] [lemma = "mielipide" & deprel = "dobj"]
4.2.1.	KOHDATA	[lemma = "kohdata"]
4.2.3.	KOHDELLA & VAKAVASTI	[lemma = "kohdella"] [lemma = "vakavasti"]
4.3.	<i>korvan + juuressa</i>	[word = "korvan"] [word = "juuressa"]
4.3.	<i>korvien + juurissa</i>	[word = "korvien"] [word = "juurissa"]
4.2.3.	KOVA & PULA	[lemma = "kova"] [lemma = "pula"]
4.2.6.	KREATIIVINEN	[lemma = "kreatiivinen"]
4.2.3.	KRITISOIDA & ANKARASTI	[lemma = "kritisoida"] [lemma = "ankarasti"]
4.2.3.	KRITISOIDA & VAKAVASTI	[lemma = "kritisoida"] [lemma = "vakavasti"]
4.2.3.	KYSYÄ & MIELI (OBJ)	[lemma = "kysyä"] [lemma = "mieli" & deprel = "dobj"]
4.2.3.	KYSYÄ & MIELIPIDE (OBJ)	[lemma = "kysyä"] [lemma = "mielipide" & deprel = "dobj"]
4.2.3.	KÄSITELLÄ & VAKAVASTI	[lemma = "käsitellä"] [lemma = "vakavasti"]
4.2.3.	KÄYDÄ & HERMO (ALL)	[lemma = "käydä"] [msd = ".*All.*" & lemma = "hermo"]
4.2.4.	LAHJOITTAA & OBJ	[deprel = "dobj"] [lemma = "lahjoittaa"]
4.2.4.	LAHJOITTAA & VAPA* (OBJ)	[lemma = "vapa.*" & deprel = "dobj"] [lemma = "lahjoittaa"]
4.4.4.	<i>laskien + yhteen</i>	[word = "laskien"] [word = "yhteen"]
4.5.1.	LUOTTAA + [0–1 sana] + ETTÄ	[lemma = "luottaa"] [] {0,1} [lemma = "että"]

Luku	Hakusana tai -yhdistelmä	CQP-kyselylauseke
4.5.1.	LUOTTAA + SIIHEN + [0–1 sana] + ETTÄ	[lemma = "luottaa"] [word = "siihen"] [] {0,1} [lemma = "että"]
4.2.6.	LUOVA	[lemma = "luova"]
4.2.3.	LÄHTEÄ & ELÄKE (ALL)	[lemma = "lähteä"] [msd = ".*All.*" & lemma = "eläke"]
4.2.3.	MENNÄ & HERMO (ALL)	[lemma = "mennä"] [msd = ".*All.*" & lemma = "hermo"]
4.5.6.	<i>Mikä + koskee</i>	[word = "Mikä"] [word = "koskee"]
4.4.1.	MINIMAALINEN & MÄÄRÄ	[lemma = "minimaalinen"] [lemma = "määrä"]
4.4.1.	MINIMIMÄÄRÄ	[lemma = "minimimäärä"]
4.5.6.	<i>Mitä + tulee</i>	[word = "Mitä"] [word = "tulee"]
4.3.	NYKYISESTI	[lemma = "nykyisesti"]
4.3.	<i>odottamatta</i>	[word = "odottamatta"]
4.3.	ODOTTAMATTOMASTI	[lemma = "odottamattomasti"]
4.2.2.	OIKEIN + ADJ	[lemma = "oikein"] [pos = "A"]
4.2.2.	OIKEIN + TÄRKEÄ	[lemma = "oikein"] [lemma = "tärkeä"]
4.2.2.	<i>ole</i> + [0–1 sana] + ERITTÄIN + ADJ	[word = "ole"] [] {0,1} [lemma = "erittäin"] [pos = "A"]
4.2.2.	<i>ole</i> + [0–1 sana] + HYVIN + ADJ	[word = "ole"] [] {0,1} [lemma = "hyvin"] [pos = "A"]
4.2.2.	<i>ole</i> + [0–1 sana] + KOVIN + ADJ	[word = "ole"] [] {0,1} [lemma = "kovin"] [pos = "A"]
4.2.2.	<i>olen, olet, on, olemme, olette, ovat, ollaan</i> + [0–1 sana] + ERITTÄIN + ADJ	[(word = "olen" word = "olet" word = "on" word = "olemmme" word = "olette" word = "ovat" word = "ollaan")] [] {0,1} [lemma = "erittäin"] [pos = "A"]

Luku	Hakusana tai -yhdistelmä	CQP-kyselylauseke
4.2.2.	<i>olen, olet, on, olemme, olette, ovat, ollaan</i> + [0–1 sana] + HYVIN + TÄRKEÄ	[(word = "olen" word = "olet" word = "on" word = "olemmme" word = "olette" word = "ovat" word = "ollaan")] [] {0,1} [lemma = "hyvin"] [lemma = "tärkeä"]
4.2.2.	<i>olen, olet, on, olemme, olette, ovat, ollaan</i> + [0–1 sana] + HYVIN + ADJ	[(word = "olen" word = "olet" word = "on" word = "olemmme" word = "olette" word = "ovat" word = "ollaan")] [] {0,1} [lemma = "hyvin"] [pos = "A"]
4.2.2.	<i>olen, olet, on, olemme, olette, ovat, ollaan</i> + [0–1 sana] + KOVIN + ADJ	[(word = "olen" word = "olet" word = "on" word = "olemmme" word = "olette" word = "ovat" word = "ollaan")] [] {0,1} [lemma = "kovin"] [pos = "A"]
4.2.2.	<i>olen, olet, on, olemme, olette, ovat, ollaan</i> + [0–1 sana] + KOVIN + TÄRKEÄ	[(word = "olen" word = "olet" word = "on" word = "olemmme" word = "olette" word = "ovat" word = "ollaan")] [] {0,1} [lemma = "kovin"] [lemma = "tärkeä"]
4.2.2.	OLLA + [0–1 sana] + HYVIN + MIELENKIINTOINEN	[lemma = "olla"] [] {0,1} [lemma = "hyvin"] [lemma = "mielenkiintoinen"]
4.2.2.	OLLA + [0–1 sana] + KOVIN + MIELENKIINTOINEN	[lemma = "olla"] [] {0,1} [lemma = "kovin"] [lemma = "mielenkiintoinen"]
4.2.4.	OSALLISTUA	[lemma = "osallistua"]
4.2.1.	OSATA & PUHUA	[lemma = "osata"] [lemma = "puhua"]
4.2.1.	OSATA (AKT)	[msd = ".*Act.*" & lemma = "osata"]
4.2.3.	OTTAÄ & VAKAVASTI	[lemma = "ottaa"] [lemma = "vakavasti"]
4.2.3.	OTTAÄ & VALOKUVA (OBJ)	[lemma = "ottaa"] [lemma = "valokuva" & deprel = "dobj"]
4.3.	PAISTE	[lemma = "paiste"]
4.3.	PAISTE (INE)	[msd = ".*Ine.*" & lemma = "paiste"]

Luku	Hakusana tai -yhdistelmä	CQP-kyselylauseke
4.2.3.	PALATA & ASIA (ILL)	[lemma = "palata"] [msd = ".*Ill.*" & lemma = "asia"]
4.2.3.	PALATA & SEIKKA (ILL)	[lemma = "palata"] [msd = ".*Ill.*" & lemma = "seikka"]
4.2.3.	PARANTAA & MAHDOLLISUUS (OBJ)	[lemma = "parantaa"] [lemma = "mahdollisuus" & deprel = "dobj"]
4.2.4.	PILAANTUA	[lemma = "pilaantua"]
4.2.4.	PILAANTUA & TILANNE (SUBJ)	[lemma = "tilanne" & deprel = "subj"] [lemma = "pilaantua"]
4.2.3.	PUHUA & MIELIPIDE (OBJ)	[lemma = "puhua"] [lemma = "mielipide" & deprel = "dobj"]
4.2.6.	RELATIIVINEN	[lemma = "relatiivinen"]
4.2.6.	RESONANSSI	[lemma = "resonanssi"]
4.3.	<i>riittävän</i> + ADV	[word = "riittävän"] [pos = "Adv"]
4.3.	RIITTÄVÄSTI + ADV	[lemma = "riittävästi"] [pos = "Adv"]
4.2.2.	SANGEN + MUU	[lemma = "sangen"] [lemma = "muu"]
4.2.3.	SANOJA & MIELIPIDE (OBJ)	[lemma = "sanoja"] [lemma = "mielipide" & deprel = "dobj"]
4.3.	SELLAISESTI	[lemma = "sellaisesti"]
4.2.6.	SIIS (VIRKKEEN ALUSSA)	[lemma = "siis" & lbound(sentence)]
4.2.4.	SISÄLTÄÄ	[lemma = "sisältyä"]
4.2.6.	SUHTEELLINEN	[lemma = "suhteellinen"]
4.3.	<i>suhteellisen</i> + ADV	[word = "suhteellisen"] [pos = "Adv"]
4.3.	SUHTEELLISESTI + ADV	[lemma = "suhteellisesti"] [pos = "Adv"]
4.2.3.	SUJUA & HUONOSTI	[lemma = "sujua"] [lemma = "huonosti"]
4.2.3.	SUJUA & VÄÄRIN	[lemma = "sujua"] [lemma = "väärin"]

Luku	Hakusana tai -yhdistelmä	CQP-kyselylauseke
4.2.3.	SUURENTAA & MAHDOLLISUUS (OBJ)	[lemma = "suurentaa"] [lemma = "mahdollisuus" & deprel = "dobj"]
4.3.	SÄDE	[lemma = "säde"]
4.3.	SÄDE (INE)	[msd = ".*Ine.*" & lemma = "säde"]
4.2.3.	SÄÄ & KANNUSTAVA	[lemma = "sää"] [lemma = "kannustava"]
4.2.3.	SÄÄ & SUOTUISA	[lemma = "sää"] [lemma = "suotuisa"]
4.5.2.	TARVITA & TAPPAA (-MINEN, OBJ)	[msd = ".*minen.*" & lemma = "tappaa"] [lemma = "tarvita"]
4.5.2.	TARVITA + [0–1 sana] + -MINEN (OBJ)	[msd = ".*minen.*" & deprel = "dobj"] [lemma = "tarvita"]
4.2.1.	TAVATA	[lemma = "tavata"]
4.2.3.	TEHDÄ & VALOKUVA (OBJ)	[lemma = "tehdä"] [lemma = "valokuva" & deprel = "dobj"]
4.4.4.	<i>tehden</i> + YHTEENVETO	[word = "tehden"] [lemma = "yhteen veto"]
4.2.6.	TEKIJÄ	[lemma = "tekijä"]
4.3.	TODELLA	[lemma = "todella"]
4.2.2.	TODELLA + MIELENKIINTOINEN	[lemma = "todella"] [lemma = "mielenkiintoinen"]
4.2.2.	TODELLA + TÄRKEÄ	[lemma = "todella"] [lemma = "tärkeä"]
4.3.	TODELLISESTI	[lemma = "todellisesti"]
4.3.	TYÖ (MON ILL)	[msd = ".*Pl.*" & msd = ".*Ill.*" & lemma = "työ"]
4.3.	TYÖ (YKS ILL)	[msd = ".*Sg.*" & msd = ".*Ill.*" & lemma = "työ"]
4.2.1.	TYÖILMAPIIRI	[lemma = "työilmapiiri"]
4.2.1.	TYÖTUNNELMA	[lemma = "työtunnelma"]
4.2.3.	VAHVA & PULA	[lemma = "vahva"] [lemma = "pula"]

Luku	Hakusana tai -yhdistelmä	CQP-kyselylauseke
4.2.4.	VAIVALLOINEN	[lemma = "vaivalloinen"]
4.2.4.	VALKEA + SUBST	[lemma = "valkea"] [pos = "N"]
4.2.6.	VASTAKAIKU	[lemma = "vastakaiku"]
4.2.1.	VOIDA & PUHUA	[lemma = "voida"] [lemma = "puhua"]
4.2.1.	VOIDA (AKT IMPF)	[msd = ".*Prt.*" & msd = ".*Act.*" & lemma = "voida"]
4.4.2.	VOIMAKKAASTI & LIITTYÄ	[lemma = "voimakas"] [lemma = "liittyä"]
4.3.	VÄKI (MON)	[msd = ".*Pl.*" & lemma = "väki"]
4.3.	VÄKI (YKS)	[msd = ".*Sg.*" & lemma = "väki"]
4.4.4.	<i>yhteenvetona</i>	[word = "yhteenvetona"]
4.3.	<i>yllättäen</i>	[word = "yllättäen"]
4.3.	<i>yllättävän</i> + ADJ	[word = "yllättävän"] [pos = "A"]
4.3.	YLLÄTTÄVÄSTI	[lemma = "yllättävästi"]
4.3.	YLLÄTTÄVÄSTI + ADJ	[lemma = "yllättävästi"] [pos = "A"]